Syntactic Interactions with Information Structure in Squliq Atayal

by

(Adlay) Kun-Long Liu

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

January 2017

@ Copyright by (Adlay) Kun-Long Liu 2017. All rights reserved.
The present thesis presents my original research. All sources have been duly acknowledged.

Except where acknowledged in the customary manner, the material presented in the present thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

(Adlay) Kun-Long Liu

January 2017
Innglung na squliq ga yan ngu qsya cyux zzik na bling hogan qsya, nanak yaq baq inlungan squliq pthuyay hmaw. (Biru na Cinbaqan 20.5)

Getting information from someone can be like getting water from a deep well. If you are smart, you will draw it out. (Proverb 20.5)
Acknowledgements

“Completing a PhD is no small feat. It requires both high intellect and a great deal of tenacity.” This statement speaks every PhD student’s mind. Fortunately, I am not alone in pursuit of my dream in linguistics since I have been accompanied by a lot of people to whom I would like to express my sincere gratitude. First of all, I am deeply indebted to my supervisors, Avery D. Andrews, I Wayan Arka and Cynthia L. Allen. Dr. Andrews has been helping me a lot throughout my PhD study. I am blessed with his helps in my overseas life in Australia and in the process of thesis writing. I really enjoy being his PhD student. His humor and clear thinking make each of his supervisory meetings full of academic enjoyment. Without his open-mindedness, my creativity would not be brought into full play. Dr. Arka is always generous in sharing with me his valuable fieldwork experiences and knowledge of Austronesian languages. I greatly appreciate his helps in my thesis writing and his encouragement for the statistical analysis I employ in the present thesis. Dr. Allen provides me with the diachronic view of morphosyntax, which complements the synchronic view I learned before. I love her lectures on the History of the English Language and the time when I attended her Anglo-Saxon reading group, which gave me the chance to know the Anglo-Saxon culture and to learn Old English.

In addition, my sincere thanks also go to Johanna Rendle-Short since the statistical methods and concepts applied in the present thesis directly benefit from her lectures on the Research Design in (Applied) Linguistics. Some parts of the present thesis have
been presented at several conferences and seminars. I am indebted to the following teachers and scholars for their constructive comments, including Joan Bresnan, Mark Donohue, Wei-wen Roger Liao, Jane Simpson, Chih-hsiang Shu, Stacy Fang-ching Teng, Shu-Chuan Tseng, and Elizabeth Zeitoun. Among them, Jane Simpson, Stacy Fang-ching Teng and Elizabeth Zeitoun also gave me lots of helps in my studies.

Moreover, I would like to express my sincere thanks to the following teachers for their helps, supports and/or instruction. Although there is no space to describe how grateful to each of them I am, their names are engraved on my mind forever. They are in the alphabetical order of the surnames: Yung-O Biq, Yueh-chin Chang, Chun-Yin Doris Chen, Yuh-show Chen, Bonnie Hui-chun Chiu, Hui-chuan Jennifer Huang, Lillian M. Huang, Louise Jansen, Cherry Ing Li, Chinfa Lien, Hsueh-o Lin, Jo-Wang Lin, T.-H. Jonah Lin, Malcom Ross, Li Mei Sung, Chih-Chen Jane Tang, Ting-Chi Tang, Jen Ting, Wei-Tien Dylan Tsai, and Hsi-nan Yeh. Among them, Lillian M. Huang, Chih-Chen Jane Tang, and Wei-Tien Dylan Tsai encouraged me to conduct the research on Austronesian languages. Without their encouragement, I would not have fallen in love with the beauty of the minority languages. In addition, Chun-Yin Doris Chen, Cherry Ing Li and Malcom Ross did me a great favor in my application for postgraduate studies.

The completion of the present thesis also benefits from my informants, Yakaw Bowtu, Umas Hakaw, Masa Pisuy, Kumay Silan and Leici Mbin Tana. I have done research on the grammar of Squliq Atayal with them for more than a decade since I was a MA student. Without their generous devotions, I would not have an in-depth understanding of this beautiful language. Moreover, I sincerely thank the Ministry of Education of the Taiwanese government for offering me the scholarship covering my living costs and tuition fees in my overseas life at ANU, the ANU College of Arts and Social Sciences for sponsoring my fieldwork trips, the Australian Linguistic Society for
offering me the scholarship for the classes on the summer camp, and Taipei Municipal Shilin High School of Commerce for giving me a three-year leave for overseas studies and stable incomes for my life back in Taiwan.

At the end of my acknowledgements, I would like to express my gratitude to my family for their everlasting supports for me. They always stand behind me for each momentous decision I make in my life. I am blessed with you all.
Abstract

The present thesis investigates the syntactic interactions with information structure in Squiliq Atayal, an Austronesian language in Taiwan. Information structure (i.e. i-structure) is a representation of discourse-pragmatic information at the sentence level in accordance with addressers’ mental states in given discourse-pragmatic contexts. By observing the interactions between syntax and information structure through both quantitative and qualitative evidence, the present thesis discovers that the syntactic structures of both unmarked and marked clauses in Squiliq Atayal are formed not only for pure morphosyntactic motivations but also for discourse-pragmatic ones.

The present thesis is composed of ten chapters. Chapter 1 is a brief introduction to the ethnic background of Squiliq Atayal and the research questions of the present thesis. Chapter 2 reviews the research on information structure, topic, focus and topicality, as well as the studies on Squiliq Atayal. In addition, this chapter elaborates the methodology of the present thesis, including data sources, the orthography and our theoretical framework, Lexical-Functional Grammar (LFG).

Chapter 3 presents the basics of grammar of Squiliq Atayal, which include the rigid word order Verb-Object/Oblique-Subject, case markers and multiple voices, with related theoretical discussion on the internal structure of c-structure and the lexically determined mapping between a-structure and f-structure. In Chapter 4, the theoretical model of feature-based discourse functions is proposed. There are four discourse features: givenness, aboutness, prominence and contrast, the values of which constitute twelve discourse functions in i-structure. In addition, the proposal of one-to-many
mapping benefits the incorporation of gradience into LFG. Chapter 5 focuses on
discourse evidence from topicality and information chaining. The former studies the
topicality of subject, object, and oblique with the measurement proposed in Givón’s
works. The latter looks into how a piece of information connects two adjacent sentences.
It shows that new information is highly restricted by syntax.

Chapter 6 identifies what discourse functions in i-structure the grammatical
functions of subject, object and oblique correspond to. The conflict between quantitative
discourse evidence from topicality and qualitative syntactic evidence from
question-answer tests verifies the one-to-one general mapping between subject and
continuing topic with the application of the conversion function, which further proves
that Squiliq Atayal is a quasi-f-structural language and that the multiple voice system in
Squiliq Atayal belongs to both the role-remapping voice system and the
information-salience one in linguistic typology. Chapter 7 and 8 offer a general
description of the properties of grammaticalized topics and grammaticalized foci
respectively. Despite the complexity of the one-to-many mapping from grammatical
functions to discourse functions, it is explained by the Informational Mapping Theory
proposed in the present thesis, which directly establishes the mapping between
f-structure and i-structure.

Chapter 9 takes a diachronic perspective on the emergence of split-subjecthood in
the Austronesian languages. By comparing Squiliq Atayal and Tsou, it is hypothesized
that split-subjecthood emerged from the shift of primary continuing topic from
nominative subjects to oblique actors in NAV clauses. Chapter 10 summarizes the
contributions of the present thesis and points out some issues for further research.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>I</td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>IV</td>
</tr>
<tr>
<td><strong>TABLE OF CONTENTS</strong></td>
<td>VI</td>
</tr>
<tr>
<td><strong>SYMBOLS</strong></td>
<td>XI</td>
</tr>
<tr>
<td><strong>ABBREVIATIONS</strong></td>
<td>XII</td>
</tr>
<tr>
<td><strong>1. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 THIS INVESTIGATION</td>
<td>1</td>
</tr>
<tr>
<td>1.2 OUTLINE OF THE PRESENT THESIS</td>
<td>5</td>
</tr>
<tr>
<td><strong>2. LITERATURE REVIEW &amp; METHODOLOGY</strong></td>
<td>6</td>
</tr>
<tr>
<td>2.1 OVERVIEW</td>
<td>6</td>
</tr>
<tr>
<td>2.2 LITERATURE REVIEW</td>
<td>6</td>
</tr>
<tr>
<td>2.2.1 NOTION OF INFORMATION STRUCTURE</td>
<td>7</td>
</tr>
<tr>
<td>2.2.2 NOTION OF TOPIC &amp; FOCUS IN LFG</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3 NOTION OF TOPICALITY</td>
<td>17</td>
</tr>
<tr>
<td>2.2.4 PREVIOUS RESEARCH ON SQUIQ ATAYAL</td>
<td>18</td>
</tr>
<tr>
<td>2.3 METHODOLOGY</td>
<td>21</td>
</tr>
<tr>
<td>2.3.1 DATA COLLECTION</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1.1 Sources</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1.2 Orthography</td>
<td>27</td>
</tr>
<tr>
<td>2.3.2 THEORETICAL FRAMEWORK</td>
<td>28</td>
</tr>
</tbody>
</table>
## 2.4 SUMMARY

### 3. BASICS OF GRAMMAR

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 OVERVIEW</td>
<td>34</td>
</tr>
<tr>
<td>3.2 C-STRUCTURE</td>
<td>35</td>
</tr>
<tr>
<td>3.2.1 WORD ORDER</td>
<td>35</td>
</tr>
<tr>
<td>3.2.2 I’ OR INTERNAL S</td>
<td>39</td>
</tr>
<tr>
<td>3.2.3 SPECIFIER VS. ADJUNCT</td>
<td>48</td>
</tr>
<tr>
<td>3.3 F-STRUCTURE &amp; A-STRUCTURE</td>
<td>55</td>
</tr>
<tr>
<td>3.3.1 CASE MARKERS &amp; PRONOUNS</td>
<td>55</td>
</tr>
<tr>
<td>3.3.2 MULTIPLE VOICES</td>
<td>58</td>
</tr>
<tr>
<td>3.3.3 GRAMMATICAL FUNCTIONS</td>
<td>62</td>
</tr>
<tr>
<td>3.3.4 LEXICALLY DETERMINED MAPPING</td>
<td>72</td>
</tr>
<tr>
<td>3.4 SUMMARY</td>
<td>78</td>
</tr>
</tbody>
</table>

### 4. I-STRUCTURE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 OVERVIEW</td>
<td>80</td>
</tr>
<tr>
<td>4.2 PRIMITIVES</td>
<td>80</td>
</tr>
<tr>
<td>4.2.1 DEFINITION</td>
<td>81</td>
</tr>
<tr>
<td>4.2.2 DISCOURSE FUNCTIONS &amp; FEATURES</td>
<td>82</td>
</tr>
<tr>
<td>4.2.3 I-STRUCTURE CONSTRAINTS</td>
<td>93</td>
</tr>
<tr>
<td>4.3 GENERAL MAPPING</td>
<td>95</td>
</tr>
<tr>
<td>4.3.1 ONE-TO-MANY MAPPING</td>
<td>95</td>
</tr>
<tr>
<td>4.3.2 TYPE-TOKEN RELATIONS</td>
<td>98</td>
</tr>
<tr>
<td>4.4 SUMMARY</td>
<td>101</td>
</tr>
</tbody>
</table>

### 5. DISCOURSE EVIDENCE: TOPICALITY & INFORMATION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 OVERVIEW</td>
<td>103</td>
</tr>
<tr>
<td>5.2 TOPICALITY</td>
<td>103</td>
</tr>
<tr>
<td>5.2.1 DATA</td>
<td>104</td>
</tr>
<tr>
<td>5.2.2 RESEARCH DESIGN</td>
<td>107</td>
</tr>
</tbody>
</table>
9.1 OVERVIEW 298
9.2 SPLIT-SUBJECTHOOD 298
9.2.1 NOTION 299
9.2.2 IN TSOU & SQUILIQ ATAYAL 302
9.2.2.1 In Tsou 302
9.2.2.2 In Squiliq Atayal 308
9.2.2.3 Comparisons 311
9.3 DIACHRONIC HYPOTHESIS 313
9.3.1 SYNTACTIC CHANGE 313
9.3.2 DEVELOPMENT OF SPLIT-SUBJECTHOOD 319
9.3.3 EVIDENCE 325
9.3.3.1 For CTTOP, in Tsou 326
9.3.3.2 For Shift Direction 330
9.4 FURTHER DISCUSSION 336
9.4.1 EXAMINING GF 336
9.4.2 BREAKING DEFAULT MAPPING 345
9.5 SUMMARY 349

10. CONCLUDING REMARKS 350
10.1 CONTRIBUTIONS 350
10.2 FOR FUTURE RESEARCH 351

APPENDIX 1: EFFECT OF DIFFERENT COMPONENTS 353
A1.1 MULTINOMINAL LOGISTIC REGRESSION 353
A1.2 DEFICIENT MODELS 357
A1.2.1 MODEL 1, 2 & 3 357
A1.2.2 MODEL 4 359
A1.3 FULL MODEL & INTERPRETATION 360
A1.4 SUMMARY 364

REFERENCES 366
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>1(^{st}) person plural</td>
<td>Neu</td>
<td>Neutral</td>
</tr>
<tr>
<td>1S</td>
<td>1(^{st}) person singular</td>
<td>Nom</td>
<td>Nominative</td>
</tr>
<tr>
<td>2P</td>
<td>2(^{nd}) person plural</td>
<td>NonPt</td>
<td>Non-Past</td>
</tr>
<tr>
<td>2S</td>
<td>2(^{nd}) person singular</td>
<td>Obl</td>
<td>Oblique</td>
</tr>
<tr>
<td>3P</td>
<td>3(^{rd}) person plural</td>
<td>OV</td>
<td>Objective Voice</td>
</tr>
<tr>
<td>3S</td>
<td>3(^{rd}) person singular</td>
<td>Par</td>
<td>Particle</td>
</tr>
<tr>
<td>Abs</td>
<td>Absolutive</td>
<td>Perf</td>
<td>Perfective</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative</td>
<td>Pres</td>
<td>Present</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspect</td>
<td>Prog</td>
<td>Progressive</td>
</tr>
<tr>
<td>Aux</td>
<td>Auxiliary</td>
<td>Proj</td>
<td>Projective</td>
</tr>
<tr>
<td>AV</td>
<td>Actor Voice</td>
<td>Pt</td>
<td>Past</td>
</tr>
<tr>
<td>Cau</td>
<td>Cause</td>
<td>PV</td>
<td>Patient Voice</td>
</tr>
<tr>
<td>Cl</td>
<td>Classifier</td>
<td>Rea</td>
<td>Realis</td>
</tr>
<tr>
<td>Cmp</td>
<td>Complementizer</td>
<td>Rec</td>
<td>Reciprocal</td>
</tr>
<tr>
<td>Con</td>
<td>Conjunct</td>
<td>Red</td>
<td>Reduplication</td>
</tr>
<tr>
<td>CV</td>
<td>Circumstantial Voice</td>
<td>Rel</td>
<td>Relativizer</td>
</tr>
<tr>
<td>Emp</td>
<td>Emphatic</td>
<td>Sub</td>
<td>Subordinator</td>
</tr>
<tr>
<td>Erg</td>
<td>Ergative</td>
<td>Top</td>
<td>Topic marker</td>
</tr>
<tr>
<td>Exc</td>
<td>Exclusive</td>
<td>Ø</td>
<td>Emptiness; Phonetically</td>
</tr>
<tr>
<td>Excl</td>
<td>Exclamation</td>
<td>*</td>
<td>Ungrammatical</td>
</tr>
<tr>
<td>Fil</td>
<td>Pause Filler</td>
<td>?</td>
<td>Marginal</td>
</tr>
<tr>
<td>Ft</td>
<td>Future</td>
<td>#</td>
<td>Wrong meaning/sense</td>
</tr>
<tr>
<td>Gen</td>
<td>Genitive</td>
<td>/</td>
<td>Or</td>
</tr>
<tr>
<td>Hab</td>
<td>Habitual</td>
<td>-</td>
<td>Morphologic boundary</td>
</tr>
<tr>
<td>Imp</td>
<td>Imperative</td>
<td>=</td>
<td>Clitic boundary</td>
</tr>
<tr>
<td>Impf</td>
<td>Imperfective</td>
<td>&lt; &gt;</td>
<td>Infix boundary</td>
</tr>
<tr>
<td>Inc</td>
<td>Inclusive</td>
<td>[ ]</td>
<td>Modified letters/words in quotations</td>
</tr>
<tr>
<td>Irr</td>
<td>Irrealis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ld</td>
<td>Locational determiner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin</td>
<td>Linker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loc</td>
<td>Locative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td>Locative Voice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAV</td>
<td>Non-Actor Voice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABT</td>
<td>Aboutness (discourse feature)</td>
</tr>
<tr>
<td>ACT</td>
<td>ACTOR (thematic function)</td>
</tr>
<tr>
<td>ADJ</td>
<td>ADJUNCT (grammatical function)</td>
</tr>
<tr>
<td>a-structure</td>
<td>Argument structure (LFG representation)</td>
</tr>
<tr>
<td>AV</td>
<td>Actor Voice</td>
</tr>
<tr>
<td>AVM</td>
<td>Attribute-value matrix</td>
</tr>
<tr>
<td>BKINF</td>
<td>BACKGROUND INFORMATION (discourse function)</td>
</tr>
<tr>
<td>CPINF</td>
<td>COMPLETIVE INFORMATION (discourse function)</td>
</tr>
<tr>
<td>c-structure</td>
<td>Constituent structure (LFG representation)</td>
</tr>
<tr>
<td>CTTOP</td>
<td>CONTINUING TOPIC (discourse function)</td>
</tr>
<tr>
<td>CTTOP_p</td>
<td>PRIMARY CONTINUING TOPIC (discourse function)</td>
</tr>
<tr>
<td>CTTOP_θ</td>
<td>SECONDARY CONTINUING TOPIC (discourse function)</td>
</tr>
<tr>
<td>CD</td>
<td>Communicative dynamism</td>
</tr>
<tr>
<td>CG</td>
<td>Construction Grammar</td>
</tr>
<tr>
<td>CL</td>
<td>Clitics (syntactic category)</td>
</tr>
<tr>
<td>COM(P)</td>
<td>COMPLEMENT (grammatical function)</td>
</tr>
<tr>
<td>CONT</td>
<td>Contrast (discourse feature)</td>
</tr>
<tr>
<td>CSFOC</td>
<td>CONTRASTIVE FOCUS (discourse function)</td>
</tr>
<tr>
<td>CSTHM</td>
<td>CONTRASTIVE THEME (discourse function)</td>
</tr>
<tr>
<td>CSTOP</td>
<td>CONTRASTIVE TOPIC (discourse function)</td>
</tr>
<tr>
<td>CV</td>
<td>Circumstantial Voice</td>
</tr>
<tr>
<td>DF</td>
<td>DISCOURSE FUNCTION</td>
</tr>
<tr>
<td>&lt;E&gt;</td>
<td>Elicitational data</td>
</tr>
<tr>
<td>fn</td>
<td>Footnote</td>
</tr>
<tr>
<td>FOC</td>
<td>FOCUS (grammatical function)</td>
</tr>
<tr>
<td>FORM</td>
<td>PHONETIC FORM (grammatical feature)</td>
</tr>
<tr>
<td>f-structure</td>
<td>Functional structure (LFG representation)</td>
</tr>
<tr>
<td>GB</td>
<td>Government and Binding Theory</td>
</tr>
<tr>
<td>GF</td>
<td>GRAMMATICAL FUNCTION</td>
</tr>
<tr>
<td>GF̂</td>
<td>MOST PROMINENT ARGUMENT (grammatical function)</td>
</tr>
<tr>
<td>GFOC</td>
<td>Grammaticalized focus</td>
</tr>
</tbody>
</table>
GIV : Givenness (discourse feature)
GOF : GRAMMATICALIZED OVERLAY FUNCTION (grammatical function)
GTOP : Grammaticalized topic
GTOP-DG : Dangling grammaticalized topic
GTOP-LD : Left-dislocated grammaticalized topic
GTOP-LD/DG : Left-dislocated/dangling grammaticalized topic
GTOP-SUBJ : Grammaticalized topic coreferential with subjects
HLRHM : HIGHLIGHTED RHEME (discourse function)
IHRC : Internally-headed relative clause
IMT : Informational Mapping Theory
LFG : Lexical-Functional Grammar
LMT : Lexical Mapping Theory
INDEX : INDEX OF REFERENTIALITY (grammatical feature)
i-structure : Information structure (LFG representation)
ITFOC : INTRODUCTORY FOCUS (discourse function)
lit. : Literal meaning
LOC : LOCATION (thematic function)
LV : Locative Voice
MGG : Minimalist Generative Grammar
MLOGIT : Multinomial Logistic Regression
n : Note
<N> : Naturalistic data
NAV : Non-Actor Voice
OBJ : OBJECT (grammatical function)
OBL : OBLIQUE (grammatical function)
OT : Optimality Theory
PAN : Proto-Austronesian
P&P : Principles and Parameters
PAT : PATIENT (thematic function)
p. c. : Personal Communication
PIV : PIVOT (grammatical function)
Pl : Plural
POSS : POSSESSOR (grammatical function)
PRED : PREDICATE (grammatical function)
PREDFN : PREDICATE FOOTNODE (grammatical function)
PREDLINK : PREDICATE LINK (grammatical function)
PROM : Prominence (discourse feature)
PRON TYPE : PRONOUN TYPE (grammatical function)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSFOC</td>
<td>PRESENTATIONAL FOCUS (discourse function)</td>
</tr>
<tr>
<td>p-structure</td>
<td>Phonological structure (LFG representation)</td>
</tr>
<tr>
<td>PV</td>
<td>Patient Voice</td>
</tr>
<tr>
<td>QFOC</td>
<td>INTERROGATIVE FOCUS (discourse function)</td>
</tr>
<tr>
<td>RD</td>
<td>Referential Distance</td>
</tr>
<tr>
<td>SFG</td>
<td>Systematic Functional Grammar</td>
</tr>
<tr>
<td>Sg</td>
<td>Singular</td>
</tr>
<tr>
<td>s. m.</td>
<td>Slightly modified from</td>
</tr>
<tr>
<td>s-structure</td>
<td>Semantic structure (LFG representation)</td>
</tr>
<tr>
<td>SUBJ</td>
<td>SUBJECT (grammatical function)</td>
</tr>
<tr>
<td>SSTOP</td>
<td>SCENE-SETTING TOPIC (discourse function)</td>
</tr>
<tr>
<td>SWTOP</td>
<td>SWITCH TOPIC (discourse function)</td>
</tr>
<tr>
<td>TAM</td>
<td>Tense, aspect &amp; mood</td>
</tr>
<tr>
<td>TF</td>
<td>THEMATIC FUNCTION</td>
</tr>
<tr>
<td>TOP</td>
<td>TOPIC (grammatical function)</td>
</tr>
<tr>
<td>TP</td>
<td>Topic Persistence</td>
</tr>
<tr>
<td>TYPE</td>
<td>CLAUSE TYPE (grammatical function)</td>
</tr>
<tr>
<td>XCOMP</td>
<td>OPEN COMPLEMENT (grammatical function)</td>
</tr>
</tbody>
</table>
1. Introduction

Seize the moment of excited curiosity on any subject to solve your doubts; for if you let it pass, the desire may never return, and you may remain in ignorance.

--William Wart

1.1 This Investigation

Atayal is a language spoken by one of the indigenous minorities in Taiwan. There are approximately 80,000 Atayal people inhabiting in the northern mountainous areas (Lewis 2009). Because of the past language education and of the frequent and long contact with the majority Han Chinese on this small island, some of them (especially youths and children) only speak Mandarin Chinese.¹ The declining number of proficient native speakers indicates that the survival of this language has been threatened.²

Atayal genealogically belongs to the Austronesian family. The Austronesian languages in Taiwan are called Formosan languages. Although there are some debates on the specific subgrouping of the Austronesian family, Atayal together with Seediq is classified under a broader Atayalic branch descending either from the highest level of the Austronesian family—the Proto-Austronesian (PAN) (Blust 1977, 1995, 1999), or

---

¹ In the post-WWII period from 1945 to 1990s, the education in Taiwan aimed to promote Mandarin Chinese as the official language by forbidding speaking local languages, including Taiwanese (i.e. Southern Min), Hakka and indigenous languages, at school (Chou 2015).
² Sh. Chen (2006, 2010) follows the Graded Intergenerational Disruption Scale proposed in Fishman (1991) to investigate the language disruption of Atayal. It is found out that out of the three research sites, two are approaching Scale 7 and one Scale 6. Scale 8 is the severest level of language loss in Fishman (1991).
from Proto-Nuclear-Austronesian, one of the four daughters of PAN (Ross 2009).³ This language comprises two dialects, Squliq and C’uli’. Squliq Atayal is a more homogenous majority (Li 1985, 1996, Rau 2004), used as the representative dialect for Bible translation and most literary works (Li 2016). The geographic distribution of the two dialects is displayed in Figure 1-1.

Figure 1-1 Geographic Distribution of Squliq and C’uli’ (Luuva 2008)

The genealogical relations are briefly presented below (based on Li (1999) and Ross

³ For different views readers can refer to Dyen (1971), Dyen and Tsuchida (1991), Haudricourt (1965), Ross (1992), and Starosta (1995), etc. Summaries of these different views can be found in Blust (1999). Based mainly on phonological evidence, Blust (1999) argued against the existence of the Proto-Formosan language and proposed that there were ten branches immediately descending from PAN: Atayalic, East Formosan, Puyuma, Paiwan, Rukai, Tsoiuc, Bunun, Western Plains, Northwest Formosan, and Malayo-Polynesian. In this subgrouping, Formosan languages constitute nine out of ten subgroups under PAN, and Atayalic is exactly one of them. Ross (2009) adopted morphological evidence, further lumping together languages other than Puyuma, Rukai & Tsou under the branch of Nuclear Austronesian. There are a series of papers arguing about the details of this subgrouping, such as Ross (2012), Sagart (2010) and Teng & Ross (2010), and so forth. In addition, Deng & Wang (2009) argued against the diachronically linear development in historical linguistics, proposing a multiple wave expansion in the Austronesian family.
The aim of the present thesis is to investigate the interaction between syntax and information structure in Squliq Atayal. As presented later in the present thesis, it is syntactically characterized by the rigid word order verb-object-subject (VOS) and four grammatical voices. Among approximately 5,000 languages in the world, languages with S(ubject)-O(bject) word order are much more common than those with O(bject)-S(ubject) word order. That is why Greenberg’s pioneering work on linguistic typology proposed a universal fact that a nominal subject almost always preceded a nominal object and even predicted that object-before-subject languages did not occur at all or at least are excessively rare (Greenberg 1963:61). The non-existence prediction however was not born out. Tomlin (1986) found out that object-before-subject languages were about 4.23% of all the 402 languages he investigated and that VOS

---

4 This Greenberg’s universal further leads to the bipartite OV-VO typology in later typological studies such as Lehmann (1973) and Hawkins (1983).
languages occupied 2.99% of his samples. This rarity is due to the tendency that subjects usually carry given information while objects new information and the given-new discourse principle that given information tends to precede new information. Squiliq Atayal, with VOS as the unmarked word order, undoubtedly belongs to the minority, which further brings about the linguistic inquiries about how information is manifested at the sentence level in this language, what role voice alternation plays in this syntactic interaction with information structure, and what the differences between unmarked clauses with VOS and the marked clauses with a grammaticalized topic (GTOP) or a grammaticalized focus (GFOC) are.

Based on the empirical facts discovered from Squiliq Atayal, the present thesis delves into theoretical discussion on i-structure (the representation of information structure) under the framework of Lexical-Functional Grammar (LFG), including the issues on the inventory of discourse features, i-structural constraints, and the correspondences between grammatical functions and discourse functions. By examining this language through both quantitative and qualitative evidence, we will have a better understanding of how syntax and information structure work in tandem in Universal Grammar or the linguistic universe. To be more specific, the present thesis proceeds to answer the following research questions:

(1) **Empirical Questions**

- How is information structure encoded by means of word order in Squiliq Atayal?
- How is information structure associated with grammatical functions in Squiliq Atayal?  

---

5 Based on the frequency of six unmarked word order, Tomlin (1986) proposed the Theme First Principle that “in clauses information that is relatively more thematic precedes information that is less so (p.48).”

6 In the present thesis, the full form “information-structure” refers to the general concept shared in different linguistic approaches whereas the short form “i-structure” refers to the module in LFG.

7 Grammatical functions consist of subject, direct object, oblique, indirect/secondary object, etc. They are also called “grammatical relations” or “syntactic roles”. These terms are used interchangeably in the present thesis, but “grammatical functions” are used particularly in LFG-related discussion. Likewise, the linguistic term “semantic roles”, which refers to agent, actor, patient, beneficiary, and so forth, is basically interchangeable with “thematic relations”, “thematic roles” and “thematic functions” in the present thesis.
How does information structure influence the use of voice alternation in Squilq Atayal?
How does information structure affect marked syntactic structures, such as topicalization or pseudo-clefting?

**Theoretical Questions**

- What discourse functions does i-structure have?
- How does i-structure correspond with other modules, such as f(unctional)-structure and c(onstituent)-structure?

### 1.2 Outline of the Present Thesis

The present thesis falls into four parts. The first part, comprising Chapter 1 and Chapter 2, is an introduction to the target language, previous research and methodology. The second part, including Chapter 3, Chapter 4 and Chapter 5, presents empirical facts and basic theoretical issues from syntax and discourse not only to provide an in-depth investigation on information coding in Squilq Atayal but also to lay the groundwork for advanced theoretical discussion. The third part is made of Chapter 6, Chapter 7 and Chapter 8. They investigate the interactions between i-structure and other modules in unmarked clauses and marked clauses like topicalization and pseudo-cleft. The last part is composed of Chapter 9 and Chapter 10, the former of which offers a hypothesis about the emergence of split-subjecthood phenomena based on the interactions between i-structure and other syntactic modules and the latter of which is the conclusion of the whole thesis.

but “thematic functions” are used particularly in LFG-related discussion.
2. 

Literature Review & Methodology

*I write for no other purpose than to add to the beauty that now belongs to me.*

---Jack London

2.1 Overview

This chapter focuses on two key elements of the present study: one is what academic achievements previous research (particularly the studies under the LFG framework) has made, and the other is the research methodology employed in the present thesis. It is shown in the first half of this chapter that the study of information structure as well as relevant notions, such as topic, focus and topicality, is rooted in forerunners’ legacy in linguistics. In addition, some key empirical facts on information structure in Squilq Atayal have also been touched upon in previous works. The second half of this chapter will sketch the theoretical framework adopted by the present thesis and explain the method of data collection and orthography.

2.2 Literature Review

This section briefly reviews previous studies on several information-structural notions, including information structure itself, topic, focus, and topicality, and on Squilq Atayal. Since there are a large number of previous studies, the presentation in this section mainly focuses on those in the framework of LFG.
2.2.1 Notion of Information Structure

Deeply rooted in the functionalist view that syntactic structures are the results of communication and that “a generalization about grammatical patterning might be attributed to the most orderly or efficient means of conveying information (Newmeyer 1998:10)”, Halliday (1967b) proposed the linguistic term information structure for the first time in the history of linguistics:8,9

Any text in spoken English is organized into what may be called ‘information units’. The distribution of the discourse into information units is obligatory in the sense that the text must consist of a sequence of such units. But it is optional in the sense that the speaker is free to decide where each information unit begins and ends, and how it is organized internally; this is not determined for him by the constituent structure. Rather could it be said that the distribution of information specifies a distinct constituent structure on a different plane; this ‘information structure’ is then mapped onto the constituent structure as specified in terms of sentences, clauses and so forth, neither determining the other. (Halliday 1967b:200)

In Halliday’s framework, information structure was parallel to the clausal structural unit and was instantiated as tone groups in phonology (Halliday 1985, 1994, Halliday & Matthiessen 2004). It was made up of two functions: the Given (recoverable information) and the New (unrecoverable information), which were closely associated with topic and focus.10

---

8 The academic chasm between the functionalist view and the formalist view is the extension of the bifurcation between two general academic thoughts: empiricism and rationalism. According to Robins (1967), the debate between empiricists and rationalists started from the 16th century. The empiricists claimed that “all human knowledge is derived from sense impressions and the operations of the mind upon them in abstraction and generalization (Robins 1967:112)” while “the rationalists sought for the certainty of knowledge…in the irrefutable truths of human reason (Robins 1967:112)”. These two thoughts are not completely incompatible with one another, and some linguistic approaches have been endeavoring to strike a mean, like cognitive linguists and LFG syntacticians (Falk 2001, Lakoff 1987).

9 There are also other labels used by linguists, such as “information packaging” in Chafe (1976) and “informatics” in Vallduví (1992).

10 It is easy to mistake the topic-comment terminology with another terminology theme-rheme in System-Functional Grammar. As elaborated in Halliday (1994), topic is just “one particular kind of theme” (p.38), which further includes conjuncts and different kinds of adverbials.
Although Halliday proposed the notion of information structure, he did not give a precise definition. The most influential definition of information structure in linguistics was proposed by Lambrecht (1994) under the framework of Construction Grammar, which believed that “discourse function is inherent in grammatical form (Lambrecht 1994:339)”:

**INFORMATION STRUCTURE:** That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts. (Lambrecht 1994:5)

Lambrecht’s definition mentioned two bases of information structure: first, the mental states of interlocutors provided its cognitive basis, and second, the interpretation in given discourse contexts provided its discourse-pragmatic basis. The mental state of interlocutors and discourse context are fundamental to research on information structure, usually formalized as different features in linguistic theory.11

How information structure is realized in linguistic theory differs from framework to framework. For instance, Minimalist Generative Grammar (MGG), under the framework of which plenty of works on information structure have lately been published, does not treat information structure as an independent module; put differently, no particular component in this framework is canonically termed information structure.12 Information-structural notions, such as topic and focus, are realized as

---

11 The use of the term “formalize/formalization” should not be viewed as formalism in linguistics. Formalization is not the exclusive property of formalism since some functionalist approaches, such as Systematic-Functional Grammar, Role and Reference Grammar and Construction Grammar, also formalize language into concrete units. The reason is as Halliday (1974) said: the traditional (system-structure) framework is “insufficiently explicit, and needs to be underpinned by a formalized generative model (p. 45)”. Moreover, as mentioned in Newmeyer (1998), formalism is often misunderstood because of the ambiguous meaning of “formal”, which can be interpreted in the sense of “pertaining to grammatical form and opposing meanings as well as uses.” In fact, semantic meaning does play a role in the formalist approaches.

12 Chomsky (1965) pointed out the possible analysis suggested by Paul Kiparsky that “Topic-Comment is the basic grammatical relation of surface structure corresponding (roughly) to the fundamental
features interpreted at Logic Form, and their intonational properties are realized at Phonetic Form (Radford 2009). Syntactic elements carrying these features must move to or be base-generated at the traditional CP at the left periphery of a tree diagram, which is split into several different functional projections, such as Force Phrase (ForceP), Topic Phrase (TopP), Focus Phrase (FocP), and Finiteness Phrase (FinP), as proposed in Rizzi (1997).

Figure 2-1 *Rizzi’s Structure of the Left Periphery (Rizzi 1997:297)*

These different functional projections provide possible landing sites for movement.

Subject-Predicate relation of deep structure (p.221, fn.32). Outside the MGG, Fillmore (1968) claimed that subjectivization, the operation of subject selection, was “primary topicalization”, one kind of topicalization. This claim was close to Paul Kiparsky’s idea although there was a great difference on the application of transformation.

According to Schwabe & Winkler (2007), the pioneer studies postulating information-structural features are Chomsky (1971, 1976) and Jackendoff (1972).

Where these information-structural features come from is controversial in MGG, which assumes that words are picked out from Lexicon before merger, since it is impossible to postulate that words carry these features in Lexicon (Erteschik-Shir 2007). That is why some MGG syntacticians, such as Erteschik-Shir (1997, 1999) and López (2009), advocated an independent level of information structure (or Σ-structure in Zubizarreta (1998) or focus structure in Erteschik-Shir (1997, 1999)), but their proposals were not widely accepted. In Zubizarreta (1998) and Erteschik-Shir (1997, 1999), information/focus/Σ structure is located in between syntactic structure and PF/LF representation, still inside the scope of Chomsky’s syntactic computation; nevertheless, López (2009) places information structure outside the Chomsky’s computational system, as an input to discourse.

Aissen (1992) already postulated a similar but rough structure, in which topic constructions are syntactically divided into external topics and internal topics, and semantically the former corresponds to new topics and the latter to the continuing topics. Aissen didn’t clearly specify the syntactic positions, but it appeared that external topics are phrases outside CP, internal topics are phrases inside CP and that focus phrases are outside IP.

In the cartographic approach, initially proposed by Italian syntacticians like Luigi Rizzi and Guglielmo
Movement is an important means for MGG to show the influence of information upon syntax. The above phrases are present in a structure only when a constituent bears related formal features to be sanctioned by a Spec-head criterion; on the other hand, if there exists no such constituent in a structure, these projections will be syncretized into a single projection corresponding to the traditional CP (Rizzi 1997).

Counter to MGG, the theoretical framework adopted in the present thesis, LFG, is a parallel theoretical framework treating information structure as an independent representation (called i-structure) existing along with several other grammatical representations, such as functional-structure (f-structure), argument structure (a-structure), and constituent structure (c-structure). Despite the parallelism of LFG, the incorporation of information structure into this framework had not emerged until King’s work in 1990s.

In the LFG tradition, topic and focus were viewed as grammatical functions in f-structure (i.e. TOP(IC) and FOC(US)) that expressed relations related to discourse representation, termed as “grammaticalized/grammaticized/syntacticized discourse functions” (Bresnan 2001). This treatment was faced with the challenge from the cases of predicate focus. As King (1995, 1997) pointed out, verb focus, which covered only the verb itself, and predicate focus, which covered both the verb and its object, could not be adequately represented in f-structure because the grammatical function FOC could only be associated with the whole attribute-value matrix (AVM) composed of subject, predicate and object but not with part of it. Therefore, King (1997) argued for the

Cinque, these functional projections are jointly regarded as the information layer/domain, to some extent equivalent to the information structure in other theories (Carnie 2010, Cinque & Rizzi 2008). Studies like Belletti (2004), Cognola (2010), Drubig (2007), É. Kiss, Katalin (2014), Hinterhölzl (2006), Hsu (2008), Jayaseelan (2001, 2008), Ndayiragwe (1999), Paul (2002, 2005), Poletto (2006, 2014), and Tsai (2015), among others, further expanded Rizzi’s Split-CP Hypothesis with evidence from word order of both European and non-European languages to claim that information-structural positions occurred both at the left periphery of the clausal phrase and at the left periphery of the smaller vP. This proposal was called the parallel-phases hypothesis, which could make their theoretical framework more economical in that it was faster to acquire different phases of similar configurations in the process of language acquisition.

Some MGG studies, such as Abels (2012), Neeleman & Vermeulen (2012) and Reeve (2010, 2011, 2012), argue against the various functional projections in the cartographic approach, and they propose that information-structural phenomena can be solved by existing mechanisms like Relativized Minimality.
existence of a distinct representation called i(nformation)-structure to accommodate information-structural notions and to solve the problems of the mismatch between them and f-structure. King’s model of grammar is illustrated below:

Figure 2-2 King’s LFG Model (King 1997:8)

Choi (1997, 1999, 2001) made another landmark on the study of the syntactic interaction with i-structure in LFG. First of all, the primitives of i-structures were refined. Four discourse functions were distinguished with two partitioning features [±New] and [±Prom]. The specific details are left to Section 2.2.2. Second, the interaction between syntax and information structure was clearly conceptualized through constraint ranking in the OT-LFG framework, which placed syntactic requirements, such as canonical word order, and information-structural requirements, such as the anti-focality effect that a scrambled element must necessarily be unfocused, on the same panel. The details of the OT-LFG framework are complicated and irrelevant to the present thesis so they are sidestepped here. A brief introduction to OT-LFG can be referred to Bresnan (2000). The following figure is the inter-modular interactions in Choi (1999).
Mycock (2006, 2007) accounted for the formation of constituent question in linguistic typology. Her inter-modular interactions are illustrated below:

The inter-modular interactions were illustrated in a subtler way in Mycock’s model because c-structure did not play the only dominant role and the significance of
i-structure was enhanced.

Despite their contributions to the establishment of i-structure in LFG, the above-mentioned models generally had the following weaknesses. First, the correspondence between i-structure and f-structure was unclear because the connection between i-structure and f-structure in these models was indirect, established on their correspondence with c-structure and s-structure respectively, which countered the insight of LFG that each structure should be mutually independent but correlated at the same time. C-structure should not be an indispensable medium between i-structure and f-structure; conversely, i-structure should be linked with f-structure directly. Broadly speaking, one structure should be directly related with another without a third party as the medium. Second, it was not touched upon in detail by what rules or constraints i-structure was constrained. The formation of i-structure should not merely be a random collection of discourse functions. Third, the existence of i-structure seemed to have no bearing on unmarked clauses with canonical word order since the previous research only focused on the formation of scrambling, topicalization and constituent questions. The theoretical and practical significance of i-structure should be more than this. Put differently, i-structure should have its own status even in the formation of unmarked clauses. Last, in terms of their methodology, no discourse evidence was adopted in these studies even though discourse was supposed to lay the groundwork for the analysis of i-structure, the better model of which should implement evidence from both syntax and discourse.

2.2.2 Notion of Topic & Focus in LFG

The study on information-structural notions, such as topic and focus started much longer before the linguistic term information structure was proposed in 1967. Before the
advent of modern linguistics, they were touched upon by some medieval grammarians. In modern linguistics, according to Firbas (1974), the notion of topic/theme originated from the “point of departure” of a sentence in the French scholar Henri Weil’s book De l’ordre des mots dans les langues anciennes comparées aux langues modernes (The Order of Words in the ancient languages compared with that of modern languages) published in 1844 and from the “psychological subject” in V. Ertl’s book Mluvnice česká II (Skladba) (A Grammar of Czech II (Syntax)) published in 1926. In addition, Hermann Paul’s book Prinzipien der Sprachgeschichte (Principles of the History of Language) published in 1880 already highlighted the pragmatic use of focus in question-answers pairs (Krifka & Musan 2012). Moreover, the given-new discourse principle that known information precedes new one was proposed in one of the Germanic historical linguist Otto Behaghel’s five laws at the turn of the 19th century and the 20th century (Harris & Campbell 1995).


---

18 It has even been argued by some linguists that the great work of the Roman rhetorician Marcus Fabius Quintilianus, Institutio Oratoria, already presented such notions as topic and comment (Ledgeway 2012).
19 The notion of focus originated from Halliday (1967b), which stated that “[information focus] involves the selection, within each information unit, of a certain element or elements as points of prominence within the message. (p.203)”
frameworks and the lack of their mutual compatibility.20

The gestation of information-structural notions in LFG started from Bresnan & Mchombo (1987), which analyzed TOPIC as an independent grammatical function in f-structure based on the evidence of Chichewa. Since then, information-structural notions officially entered this framework. As mentioned in Section 2.2.1, because of King’s proposal, they were relocated in i-structure in later LFG works, termed as discourse functions. In addition to TOP (topic), more discourse functions were proposed, such as FOC (focus), BCK (background) and COMP.INF (completive information)(Butt & King 2000, King 1997). Choi (1997, 1999, 2001) refined the inventory of discourse functions with two partitioning discourse features [±New] and [±Prom] to distinguish four information types: TOP (topic), CONT.FOC (contrastive focus), TAIL (tail), and COMP.FOC (completive focus).21 [±New] stood for newness, differentiating whether or not the information was new to the addressee and [±Prom] stood for prominence, differentiating whether or not the information was highlighted by means of prosody-marking or morphosyntax-marking and stood out in the foreground. TOP (topic) was defined as non-new and prominent information. CONT.FOC was defined as new and prominent information. TAIL was defined as non-new and nonprominent information. COMP.FOC was defined as new and nonprominent information. The classification is illustrated in the following table from Choi (2001:148):

---

20 Schlobinski & Schütze-Coburn (1992) and Szwedek (1990) even suggested that the notion of topic should be abandoned because of its inconsistent definitions in different works.

21 In linguistic literature, there are other terms for these notions. Contrastive focus is also termed as “narrow focus”, “identificational focus”, or “operator focus”. Tail is also termed as “background information”. Completive focus is also termed as “wide focus”, “information focus” or “rHEME”.
There are two strengths on Choi’s feature-based information structure. First, it enhances the accuracy of the definition of each discourse function as long as discourse features are clearly defined. Second, as pointed out by Choi (2001:92), “[o]ne of the advantages of this feature-based information structure is that it can crossrefer to more than one distinct information type.” This advantage will be manifested in the mapping theory proposed later in the present thesis.

Choi’s classification however has its own limitations since cross-linguistically, more than four discourse functions have been identified, and there is no doubt that more discourse features are needed to differentiate them. For instance, there are languages, like Gawwada (an African language) and Finnish, which syntactically mark contrastive entities without making any differentiation on whether their information is given or new (Konietzko & Winkler 2010, Molnár & Winkler 2010, Payne 1995, Tosco 2010). It has been proposed by some linguists that contrast (or “kontrast”) should exist as an autonomous information-structural notion (Molnár 2002, Neeleman & Vermeulen 2012, Vallduví & Vilkuna 1998).22 The distinction between contrast and newness has also been proved in phonetic research (Katz & Selkirk 2011). It follows that the two partitioning discourse features [±New] and [±Prom] do not suffice to explain diverse linguistic phenomena.

The present thesis will preserve the strength of Choi’s insight, that is, the use of

<table>
<thead>
<tr>
<th>[+Prom]</th>
<th>[−Prom]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[−New]</td>
<td>Topic, Link</td>
</tr>
<tr>
<td>[+New]</td>
<td>Contrastive Focus</td>
</tr>
</tbody>
</table>

---

22 Some linguists proposed that contrastive topic was a composite of topic and focus rather than a topic plus contrast (Buring 1997, Erteschik-Shir 2007, Krifka 2007), while some argued against the independent existence of contrast (Repp 2010).
discourse features, and expand the inventory of discourse features to accommodate more discourse functions in i-structure in order to account for the empirical facts in Squliq Atayal. One may argue against this feature-based approach since linguists can never exhaust the possibility of postulating new discourse features.\textsuperscript{23} As long as we are willing, we can classify information to the extent that every piece of information itself is an independent category with its own unique feature. In other words, the number of both discourse features and functions seems infinite. This worry nevertheless is unnecessary because linguists only focus on the features and functions having a bearing on linguistic research.

2.2.3 Notion of Topicality

As shown in lots of studies on information structure, such as Chafe (1976) and Givón (1977, 1983b, 1990, 1994), discourse always follows a global discourse topic, and every nominal in the discourse carries different degree of topicality.\textsuperscript{24} Topicality is the property of nominal participants based on the extent of the information being about them (Givón 1993). The more topical a nominal is, the more accessible and the more important information it conveys. No matter what kind of information a nominal carries, the discourse topic is passed down from one clause to another, forming topic continuity. In other words, every nominal plays a role in the topic continuity. Topic continuity in discourse has a great impact on syntactic structures. For languages with flexible word

\textsuperscript{23} Palmer (1981) and Lyons (1995) criticized the means of componential analysis, i.e. the feature-based approach adopted here. What they criticized is its application to lexical semantics, and most of the deficiencies they mentioned do not occur in our analysis. The problems about which sense is more basic, about the disability of composition of several components, about the non-contradiction between two senses, and about the discrepancy between physical reality and psychological reality all have nothing to do with the proposal here. Even in semantic research, however, the application of componential analysis still makes remarkable achievements, such as the Natural Semantic Metalanguage (NSM) (Cf. Wierzbicka 1972, 1996, inter alia).

\textsuperscript{24} Another similar relative concept is focality. Dik (1997) proposed that “the focal information in a linguistic expression is that information which is relatively the most important or salient in the given communicative setting, and considered by S [the speaker] to be most essential for A [the addressee] to integrate into his pragmatic information (p.326).” Note that focal information in Dik’s work did not equal to new information even though sometimes these two notions were overlapped on one expression. There were some terms of similar concept in linguistic literature, such as “foci” in Chafe (1976, 1979) and “dominance” in Erteschik-Shir & Lappin (1979) and Erteschik-Shir (1988).
order, such as Ute, its word order is determined by the discourse-pragmatic factor, i.e. whether the information carried by each nominal is given or new (Givón 1983a). For languages with rigid word order, like English, different syntactic strategies are adopted to arrange information, such as syntactic coding, passivization and topicalization (Givón 1993).

Topicality is successfully and widely applied to quantitative research on minority languages, and most of these studies adopt Givón’s method, such as Brainard (1994), Cooreman (1983), Pastika (1999), and Payne (1994), to name a few on Austronesian languages. Givón’s statistical measurement includes referential distance (RD), also called “look-back”, and topic persistence (TP), also called “decay”. The former assesses the gap between the previous occurrence of discourse of a referent and its current occurrence in a clause. The shorter the RD is, the more recent information this occurrence conveys, while the longer the RD is, the more remote information this occurrence conveys. The latter counts the number of times the referent recurs within the next 10 clauses following its present occurrence. The more the TP is, the more important information this referent carries, while the less the TP is, even zero, the less information this referent carries. This method is applied in the present thesis to provide statistical evidence for the correspondences between grammatical functions in f-structure and discourse functions in i-structure.

2.2.4 Previous Research on Squliq Atayal

In their sketch of Squliq Atayal grammar, Ogawa & Asai (1935) pointed out the existence of ga (later known as topic marker), identifying it as a marker linking subject and predicate, like ga and wa in Japanese. Egerod (1966, 1969, 1980) viewed ga as

---

25 The difference between the nominative case ga and the topic marker wa in Japanese was not known at that time since this identification was first made by Kuno (1972). According to Iori (2005[2001]), it was rarely known that traditional Japanese grammarian Daizaburou Matsushita had proposed similar distinction in the early 20th century.
the particle of precentral exposure. Information-structural notions had not been touched upon until Rau (1992) identified \( ga \) as a topic marker and claimed that subject conveyed old information while predicate conveyed new information. Despite her empiricism-based proposal and morphosyntax-orientation, Rau (1992) indeed opened the window to the study on the interaction between syntax and information structure in Squiq Atayal. Egerod (1993) further compared \( ga \) with \( qu' \), showing that the former indicated topic/comment sentence type while the latter subject/predicate sentence type.

L. Huang (1994) was a pioneering work of adopting statistical evidence even though the statistical part only played a minor role in the whole paper. She counted the frequency of actor voice (AV) and non-actor voice (NAV) in transitive clauses of one story, finding out that the ratio was 30:52. Thus, it was claimed that NAV clauses were actually the canonical transitive clauses while AV clauses with two arguments should be regarded as antipassive clauses. This paper only provided the basic descriptive frequency without further proving whether the ratio was statistically significant or not.

Adopting the notion of transitivity in Hopper & Thompson (1980), Rau (1994) analyzed the story line in a folktale and claimed that “non-referential or indefinite O’s show correlation with the verb morphology, case marking, and word order characteristics of ‘intransitive’ clauses in Atayal (p.509)”. That is to say, the clauses with a non-referential or indefinite patient might be intransitive to some extent. This idea was further explored in Y. Chen (2007). Under the same hypothesis of transitivity, Y. Chen (2007) studied the correlation between transitivity and foregrounding/backgrounding with the assistance of statistics. One of her findings was that word order might affect transitivity: the word order verb-patient-actor had an indefinite patient while the word order verb-actor-patient had a definite patient. Despite the close relationship between definiteness and information types, these two studies did not really go into the discussion on information structure.

Rau (2000) and S. Huang & Tanangkingsing (2011) pushed forward our
understanding of topicality in Squliq Atayal by adopting inferential statistics, which had the power to infer the overall fact from limited samples. Rau (2000) made use of VARBRUL, the linguistic software designed for binary logistic regression, to study the factors behind the variation of preverbal and postverbal subjects. She discovered that postverbal subjects were associated with old information and topic continuity while preverbal subjects were with new information and topic discontinuity: “[R]ecent second mentions, resumptives, simple NPs, bound personal pronouns, zeroes, and interrogative pronouns favor the VS order whereas text-structural boundary, first mentions, modified NPs, free pronouns, and demonstrative pronouns favor the SV order (Rau 2000:227).”

Nevertheless, the scope of her study was confined with VARBRUL, which only allowed a dependent variable of two categorical values—that is, preverbal subject or postverbal subject. It did not suffice for the demonstration of the overall information-structural properties of different grammatical functions. In addition, the so-called “preverbal subject” was a misnomer because preverbal subjects in her study were grammaticalized topics (GTOPs) in reality, the properties of which will be presented later in the present thesis.

S. Huang & Tanangkingsing (2011) was a study more faithfully following Givón’s method, and it aimed to make a comparison of the transitivity phenomenon in three Formosan languages, Kavalan, Squliq Atayal and Tsou. In their study on Squliq Atayal, the objects of AV clauses and the subjects of NAV clauses were nearly indistinguishable in terms of their definiteness and the perfectivity of these clauses. The essential difference between them was their topic persistence: the former had lower topic persistence while the latter had higher topic persistence. The assumption behind their

---

26 Logistic regression is a statistical method to test whether a categorical dependent variable is influenced by the (either categorical or continuous) independent variables the researcher assumes. Binary logistic regression only permits a dependent variable of two categorical values (Field 2009, Gray & Kinnear 2012, Pallant 2007, Yang 2004).

27 “Resumptives” in Rau’s study indicated subjects reintroduced into discourse after a lapse (Rau 2000:219).
study was that the AV transitive clauses were extended intransitive clauses in nature, wherein there were oblique-marked NPs (i.e. the AV objects in the present thesis), whereas the NAV clauses were canonical transitive clauses. Put differently, they compared the patients of these two types of clauses to examine which one had higher transitivity--that is to say, if it did, it was a real transitive clause. In addition to a small methodological flaw that this paper adopted a loose one-tailed probability, its focus was confined on the topicality of patient, without presenting a general picture of how information was manifested in a sentence.

In the lately published introductory grammar of Squiq Atayal, L. Huang & Hayung (2016), it was briefly discussed that topic sentences aimed to prepose known information to the sentence-initial position usually for the sake of contrast and comparison.

The abovementioned studies all broadened our knowledge of the information-structural notions in Squiq Atayal. Nevertheless, both their foci and methods were quite different from each other so that they either individually or as a whole did not delineate a general picture of how information structure was constituted and how it really interacted with syntax.

### 2.3 Methodology

This section discusses the research methods of the present thesis from two aspects: data collection and theoretical framework. The former is about the sources of the data and the orthographic system transcribing them. The latter gives a brief introduction to the theoretical framework adopted by the present thesis.

---

28 The idea of extended intransitive clauses comes from Dixon (1994). Extended intransitive clauses consist of intransitive verbs with two core roles, one of which is mapped onto subject while the other of which is marked by dative case or other cases, not accusative case (Dixon 1994:123).
2.3.1 Data Collection

2.3.1.1 Sources

Data collection in the present thesis follows the methodological concept of “data triangulation”: that is, “the application of more than one sampling method for data collection (Angouri 2010:34)”. In other words, data should come from more than one source. The reason behind data triangulation is simple: to avoid the weakness of different methods, which may cause the partial or even misleading presentation of evidence. There are three sources of data used in the present thesis: naturalistic production, elicitation, and publications, as presented in Figure 2-5:

Figure 2-5 Data Triangulation

Naturalistic production includes native speakers’ narratives, stories, speeches and daily conversations, during the collection process of which the researcher does not get involved in the production of data. This kind of data is suitable for academic research on any topic and from any perspective. Because this kind of data locate structures in a more concrete context, maximally preserving the contextual and cultural information,
linguists can always discover new linguistic phenomena which have never been studied yet. However, the frequency of some constructions is low in naturalistic production, and they do not contain any negative evidence for syntactic argumentation. To remedy the weakness of the first method, we need the second one—elicitation. Elicitational data come from the highly active involvement of linguists by asking for translation, providing grammaticality judgments, and testing manipulated structures. They bring both positive and negative evidence for syntactic argumentation, but the validity of these data always arouses some doubts because both translation and manipulation easily produce artificial data or even incorrect data. In order to achieve validity and reliability, all of (grammatical and ungrammatical) elicitational data presented in the present thesis were checked at least twice at different times in fieldwork. In addition, this method is limited in its research scope: because only few structures under investigation are tested, and the data collected with this method usually cannot be used in the research of other topics. Moreover, this method isolates structures in a context-free circumstance. It is difficult to use this method to study context-related issues. In the present thesis, the examples selected from naturalistic data will be labelled with <N> while those selected from elicitational data will be labelled with <E>.

Generally speaking, naturalistic data and elicitational data are complementary to each other. Although some field linguists strongly reject elicitational data and some theoretical linguists ignore the necessity of naturalistic data, the present thesis stands in between these two extremes, with both of them collected in the fieldwork. As mentioned in Chapter 1, Squliq Atayal is a minority language in Taiwan. Most of linguists on this language are not native speakers. For the sake of observational adequacy and descriptive adequacy, a six-month linguistic fieldwork was conducted. All of the data were recorded with digital recording appliances. The locus of the fieldwork for the present thesis is Jianshi, Hsinchu (also spelled as Xinzhu) County, Taiwan (R.O.C), as illustrated in Figure 2-6.
All informants are native speakers of Squiliq Atayal. Their biodata are listed in the following table.

Table 2-2 Biodata of Informants

<table>
<thead>
<tr>
<th>Indigenous Name</th>
<th>Chinese Name</th>
<th>Sex</th>
<th>Birth Year</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Kumay Silan</td>
<td>Tien, Zixung</td>
<td>M</td>
<td>1935</td>
<td>born &amp; living in Jianshi, able to speak Atayal, Mandarin &amp; Japanese</td>
</tr>
<tr>
<td>2. Umas Hakaw</td>
<td>Gao, Aimei</td>
<td>F</td>
<td>1950</td>
<td>born in Fuxing, living in Jianshi, able to speak Atayal &amp; Mandarin</td>
</tr>
<tr>
<td>3. Yakaw Bowtu</td>
<td>Lin, Tienyung</td>
<td>M</td>
<td>1944</td>
<td>born &amp; living in Jianshi, able to speak Atayal &amp; Mandarin</td>
</tr>
<tr>
<td>4. Leci Mbin</td>
<td>Zhu, Lizhi</td>
<td>F</td>
<td>1952</td>
<td>born &amp; living in Jianshi, able to speak Atayal, Mandarin &amp; Hakka</td>
</tr>
<tr>
<td>5. Masa Pisuy</td>
<td>Hong, Xioyu</td>
<td>F</td>
<td>1942</td>
<td>born in Fuxing, living in Jianshi, able to speak Atayal &amp; Mandarin</td>
</tr>
</tbody>
</table>

29 Bowern (2008) talks about the different connotations from the use of “informant” or “consultant”. The present thesis leaves open this terminological issue, following previous studies on Squiliq Atayal to use the term “informant”. 
All of the above informants are native speakers of Squliq Atayal with perfect mastery of this language. Because of the promotion of obligatory school education, all of them have the ability to speak the official language in Taiwan, Mandarin Chinese, making use of both Squliq Atayal and Mandarin often in their daily life.

The last data source is not from linguistic fieldwork but from published materials, which include texts recorded in previous studies by linguists, texts collected by native speakers as cultural workers, and texts of the open online archives built by linguists. The details of these published resources are presented in Table 2-3.
Table 2-3 Sources of Secondary Data

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Region</th>
<th>Text Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>collector</td>
<td>publication</td>
<td>nar./con.</td>
</tr>
<tr>
<td>1. HLM = (Huang 1993)</td>
<td>linguist</td>
<td>academic</td>
<td>Wulai</td>
</tr>
<tr>
<td>2. Egerod = (Egerod 1969)</td>
<td>linguist</td>
<td>academic</td>
<td>Wulai</td>
</tr>
<tr>
<td>4. LPJ = (Li 1996)</td>
<td>linguist</td>
<td>academic</td>
<td>Wulai</td>
</tr>
<tr>
<td>5. MT = (Huang, Huang &amp; Lin 2003a, b)</td>
<td>native cultural workers</td>
<td>cultural book</td>
<td>Fuxing</td>
</tr>
<tr>
<td>6. Y&amp;Y = (Yukih &amp; Yupas 1991)</td>
<td>native cultural workers</td>
<td>cultural book</td>
<td>Jianshi, Wufeng</td>
</tr>
<tr>
<td>7. NTNUCT = (Tohui et al. 1998)</td>
<td>written by native cultural workers</td>
<td>textbook</td>
<td>Wulai, Fuxing, Heping</td>
</tr>
<tr>
<td>8. FLA = Academia Sinica Formosan Language Archive&lt;sup&gt;32&lt;/sup&gt;</td>
<td>linguist</td>
<td>academic online corpus</td>
<td>Wulai, Jianshi (?)&lt;sup&gt;33&lt;/sup&gt;</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All examples from these resources are cited with their origins in the present thesis.

Since Squliq Atayal lacks a writing tradition, available published resources are few and the spelling is extremely inconsistent. The present thesis will normalize secondhand examples under the orthography adopted in the present thesis. However, it is because of

---

<sup>30</sup> The data from Yukih & Yupas (1991) include the preface written in Squliq Atayal but exclude the last text “The Theory and Practice of Indigenous Autonomy” since it is a translated document.

<sup>31</sup> The fifth lesson of Tohui et al. (1998) is excluded since it contains only a big table of the correspondence between tribe names and place names.


<sup>33</sup> To my knowledge, the older version of the Academia Sinica Archive mentioned that the fieldwork site was at Wulai. However, the latest version does not mention any information on the fieldwork site. Because the data collector Ms. Yu-ting Yeh is an Atayal at Jianshi, it is possible that some of the data were collected from Jianshi. Thus, both Wulai and Jianshi are listed in this table.
the lack of writing tradition that these published materials preserve the characteristics of oral language to the largest extent, without a distinct gap between written language and spoken language. That is why they are used to complement the texts collected in the fieldwork. By applying these three methods of data collection, this research will be able to achieve internal validity.

The data of sentences are from elicitation, including both grammatical and ungrammatical sentences. They are used for syntactic argumentation. Only few of them have digital sound files. Narratives are composed of 161 texts, 20 from fieldwork and 141 from publications. Only those texts from fieldwork have digital sound files. The content covers traditional legends, family histories, anecdotes, manufacturing procedures, etc. They are used for syntactic positive evidence, discourse evidence and statistical evidence. Only those firsthand texts matching the sampling criteria are used in the statistical studies (Cf. Chapter 5).

There are 14 texts of conversations, among which 12 are from previous publication and 2 are from fieldwork. The total length of conversations collected in fieldwork is 40 minutes long, including two dialogues stored as digital sound files. The data of conversations are used for syntactic positive evidence and discourse evidence.

2.3.1.2 Orthography

Atayal is a language without writing tradition. Generally speaking, phoneticians and phonologists adopt the International Phonetic Alphabet (IPA) to make a narrow transcription of sounds in this language, while linguists in other fields tend to make use of Li’s (1991) orthographic system, based on which the R.O.C. government promulgated the Writing System of Indigenous Languages in 2005. The list of the

---

34 This can be told by the prevalence of discourse markers, like *la*, *ha*, *vro*, *namu*, etc.
35 Li (2016) provides a detailed description of the development of the writing systems of the Austronesian languages in Taiwan. Squiliq Atayal is one of the languages without evoking any controversy because all official versions adopt the same set of symbols.
orthographic symbols in the present thesis is listed below:

Table 2-4 Orthography of Atayal

<table>
<thead>
<tr>
<th>Official System</th>
<th>IPA</th>
<th>Official System</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>e</td>
<td>e</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>q</td>
<td>q</td>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>'</td>
<td>?</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>c</td>
<td>ts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>β/v</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ng</td>
<td>ŋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>j</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3.2 Theoretical Framework

The theoretical framework of the present thesis, LFG, developed in the mid-to-late 70s out of the work of the syntactician Joan Bresnan and the computational linguists and psychologist Ronald M. Kaplan. LFG encodes two important dimensions inherited from previous generative studies: First, a model of grammar should contain lexical rules rather than transformational rules/movements, which are all “psychologically unrealistic (Bresnan 1978:2)”. Second, grammatical functions are not defined in terms of phrase structure but are primitives of the theory. With lexical properties and grammatical functions, there is no need for the existence of transformational rules.
For non-configurational languages, with flexible word order and complex morphology, LFG shows the fact that morphology plays a dominant role in its cooperation with phrase structure (i.e. c-structure in LFG). Take Warlpiri (a Pama-Nyungan language) for example. This language has flexible word order, but the complex case markers mark the grammatical function of each argument, as illustrated below:

(2) Kurdu-ngku wita-ngku ka maliki wajilipi-nyi. (Nash 1986:159)
child-Erg small-Erg Pres Abs.dog chase-NonPt
‘The small child is chasing the dog.’

(3) Kurdu-ngku ka maliki wajilipi-nyi wita-ngku. (Nash 1986:160)
child-Erg Pres Abs.dog chase-NonPt small-Erg
‘The small child is chasing the dog.’

The above two sentences have the same thematic relations, grammatical relations and proposition. The most substantial difference between (2) and (3) lies in the position of the adjective *wita-ngku* ‘small’, which seems to form a noun phrase with its modifiee *kurdu-ngku* ‘child’ in (2) but gets separated from its modifiee by all other elements in (3). Example (2) and (3) have the following c-structures respectively, which are adapted from Bresnan (2001:292-293).
In LFG, the sameness comes from the same lexical marking (i.e. morphology) while the difference is purely superficial c-structure (i.e. syntax) without the derivation from one to the other and without the issue about which one undergoes more steps of movement.

Words themselves pass their cases to their maximal projections (shown as \( \uparrow \text{CASE} = \text{Erg} \)), and then the maximal projections pass their grammatical functions (shown as \( \uparrow \text{Tense} = \text{Pres} \)).
(↑SUBJ) = ↓ based on the cases they get to the whole sentence. The above flat c-structures, without strict hierarchy, reflect the nonconfigurationality of Warlpiri.

As elaborated in Carnie’s (2010) discussion on the ontology of phrase structures, “tree structures [in the LFG tradition] aren’t [‘]derived[‘] per se…[they] are subject to filtering constraints (p.87).” The c-structure is constrained by other independent modules, such as f-structure, through inter-modal mapping. Take for example the above phrase structures in (4) and (5). They both are constrained by the identical f-structure (the representation of grammatical functions) shown in (6), and that is why they have the same grammatical functions:

(6)

```
PRED 'chase (SUBJ, OBJ)'
  [PRED 'child'
   SUBJ [CASE Erg
   ADJ {[PRED 'small']}]
  OBJ [PRED 'dog'
    CASE Abs
  TENSE Pres]
```

The above f-structure is comprised of AVMs, in which the attributes are grammatical functions (e.g. SUBJ and OBJ) and features (e.g. PRED, CASE, ADJ and TENSE) and the values are either atomic entities (e.g. Erg, Abs, Pres, ‘child’, etc.) or smaller AVMs. F-structures must obey the following three conditions (Falk 2001:63-64):

---

36 “Mapping” and “correspondence” are interchangeable terms in the present thesis.
37 Identical f-structure and a-structure may correspond to different c-structures. That is why non-configurational languages rely on rich morphology to reflect grammatical functions rather than on their various word orders, a fact called “Morphology competes with Syntax (Bresnan 1998, 2001).”
38 The definitions provided here are reader-friendly versions. The original definitions are in Kaplan & Bresnan (1982:181, 211-212).
(7) a. **Completeness Condition**
   All argument functions specified in the value of the PRED feature must be present in the local f-structure. All functions that receive a thematic role must have a PRED feature.

b. **Coherence Condition**
   All argument functions in an f-structure must be selected by their local PRED. Any argument function that has its own PRED feature must be assigned a thematic role.

c. **Uniqueness Condition** (also called Consistency Condition)
   Every attribute has a unique value.

The first two conditions work in tandem to stipulate that the grammatical functions selected by the predicate and the grammatical functions occurring in the f-structure of that predicate must be identical, no more and no less. The third condition constrains the representation of f-structure. These conditions further constrain the well-formedness of other modules via mapping.

The mapping between c-structure and f-structure is one of the remarkable characteristics of this framework. There are still other independent modules in LFG, among which i-structure has been discussed in the first half of this chapter. An overall picture of LFG can be attained via Bresnan (2001), Bresnan et al. (2016), Butt et al. (1999), Dalrymple (2001) and Falk (2001). Generally speaking, the most distinctive characteristic of the LFG framework is the mutual cooperation between parallel independent modules, without preference for c-structures.

**2.4 Summary**

It is shown in this chapter that because of the efforts of previous studies, the notions of information structure, topic, focus and topicality emerged in linguistics. With the need of expanding explanatory power, LFG linguists incorporated information structure into their framework, as a formal representation called i-structure parallel with other traditional representations like c-structure and f-structure. However, our
understanding of i-structure is still incomplete since its mapping with f-structure always relies on c-structure as the medium of mapping. In addition, more discourse features and functions are needed for linguists to account for empirical facts observed from different languages. As for Squliq Atayal, information-structural notions have been touched upon in previous studies but their focus is not really on information structure itself and its interactions with syntax.

In addition, the methodology of the present thesis is presented in the second half of this chapter. What is emphasized is data triangulation, that is, the use of data from multiple sources to avoid the possibility of bias to enhance the validity of the present research. At the end of it, the sketch of the framework adopted here—LFG—is provided to demonstrate the leading characteristics of this theory, such as annotated c-structure, inter-modular mapping and three conditions on the well-formedness of f-structure. The following chapters will further put more detailed, empirical flesh on these abstract and general bones mentioned above.
3. Basics of Grammar

Grammar protects us against misunderstanding the sound of an n uttered name.
--Rasenack-Huessy

3.1 Overview

Information structure cannot be studied without knowing the basics of grammar, which involves empirical facts and theoretical issues as two sides of the same coin. Based on the LFG framework, this chapter is divided into two parts: one is associated with c-structure and the other is with f-structure and a-structure. The first part will present the rigidity of word order and discuss the constituency of unmarked clauses with a detailed exploration of whether main verb and its object form a constituent or postverbal arguments form an internal S and what c-structural role oblique actor plays in a NAV clause. The second part will give a brief introduction to case markers, pronouns and multiple voices, further delving into the identification of grammatical functions and how the mapping between a-structure and f-structure works under the LFG framework.

The empirical facts presented in this section are not brand-new, with most of them already thoroughly presented or investigated in other works. Thus, in order not to overly reiterate them, the presentation of these basic grammatical systems will be as necessarily brief as possible. As for the theoretical issues, they may be discussed for other languages but not directly for Squiliq Atayal, so it is necessary to investigate them. However, their analyses will be deliberately restricted in order not to deflect readers’
attention too far from information structure.

3.2 C-Structure

This section will discuss the rigidity of the word order in Squiliq Atayal first. The rigidity leads to two essential issues on c-structure which will be investigated latter. One is whether main verbs and their objects in AV clauses form a constituent or the postverbal core arguments form an internal S; the other is what oblique actors in NAV clauses are in c-structure, adjuncts or specifiers.

3.2.1 Word Order

With regard to word order, Squiliq Atayal is a predicate-initial language, wherein predicates come first in simple clauses of unmarked word order, that is, occupying the sentence-initial position, no matter whether they are verbal or nominal. This grammatical characteristic can be demonstrated in the following three examples, the first two of which contain a verbal predicate with one argument (i.e. an intransitive verb) while the last of which contains a nominal predicate with one argument (i.e. a predicative nominal):

(8) m-qwas qu’ watan.
    AV-sing Nom Watan
    ‘Watan sings.’ <E>

(9) wagiq qu’ watan.
    tall Nom Watan
    ‘Watan is tall.’ <E>

(10) tayal qu’ watan.
    Atayal Nom Watan
    ‘Watan is an Atayal.’ <E>
In these examples, *watan* is the only argument of each predicate and it is preceded by the nominative case marker *qu*. Put differently, *watan* is the subject. In addition, all of the predicates, like the dynamic verb *mqwas* ‘sing’, the stative verb *wagiq* ‘tall’ and the predicative nominal *tayal* ‘Atayal’, occupy the sentence-initial position.\(^{39}\) It is still the same even when the predicate is a verbal complex containing more than one verbal element:

\[(11)\]
\[
\text{cyux m-qwas qu’ watan.} \\
\text{Aux.Prog AV-sing Nom Watan} \\
\text{‘Watan is singing.’ <E>}
\]

The above example contains a verbal complex composed of an auxiliary *cyux* and a main verb *mqwas* ‘sing’, and the whole verbal complex precedes the subject.\(^{40}\)

If the predicate is a transitive verb, its object must adjacently follow the verb itself, with the word order of transitive clauses being verb-object-subject (VOS).\(^{41}\) In terms of linguistic typology, Squliq Atayal is a rigid VOS language. This is illustrated below:

\[(12)\]
\[
\text{m-aniq qulih qu’ watan.} \\
\text{AV-eat fish Nom Watan} \\
\text{‘Watan eats fish.’ <E>}
\]

\(^{39}\) According to previous studies on verb classification in Atayal, such as L. Huang (2000a) and Zeitoun & Huang (2000), this language only makes a distinction between dynamic verbs and stative verbs. There is no adjective in this language. This runs afoul of Baker’s (2003) hypothesis that nouns, verbs and adjectives are three universal lexical categories.

\(^{40}\) There are two progressive auxiliaries in Squliq Atayal, *nyux* and *cyux*, with the former (the proximal progressive) indicating an action taking place close to the speaker and the latter (the distal progressive) away from the speaker (L. Huang 1993). Owing to the limitation of space and its irrelevance to the theme, this distinction will not be glossed in the present thesis.

\(^{41}\) As for what kind of clause is transitive and what argument is an object, there were different proposals in previous research. Here the traditional view in Egerod (1966), Rau (1992) and L. Huang (1993) is followed. The stance of the present thesis will be justified in Section 3.3.
In both (12) and (13), *watan is the subject, preceded by the nominative case marker *qu’. Example (12) with VOS word order is grammatical while example (13) with VSO word order is not. It is manifest that the relative order between subject and object cannot be reversed. In other words, word order determines the grammatical functions of arguments, and case marking supplements its effect but cannot override it.

Although the voice system of Squliq Atayal will be discussed later, it is necessary to state in advance that the rigidity of word order is observed not only in actor voice (AV) clauses, as seen in the above-mentioned examples, but also in non-actor voice (NAV) clauses, as exemplified below:

(14) niq-an na’ watan qu’ qulih qasa.
    eat-NAV Obl Watan Nom fish that
    ‘That fish was eaten by Watan.’ <E>

(15) *niq-an qu’ qulih qasa na’ watan.
    eat-NAV Nom fish that Obl Watan
    Intended for ‘That fish was eaten by Watan.’ <E>

In both (14) and (15), wherein the main verb *niqan ‘eat’ carries the NAV marking –*an, *qulih qasa ‘that fish’ is the subject while *watan is the oblique actor, which is preceded by the oblique case marker *na’. The difference in grammaticality lies in the relative order of subject and oblique actor. When the oblique actor *watan precedes the subject *qulih qasa ‘that fish’, the clause is grammatical, as presented in (14); conversely, when the subject precedes the oblique actor, the clause is ungrammatical, as presented in (15).

It may be claimed that the abbreviation ‘VOS’ in Squliq Atayal stands for not only ‘verb-object-subject’ in AV clauses but also ‘verb-oblique-subject’ in NAV clauses.
As a predicate-initial language, Squliq Atayal can have nominal elements in the sentence-initial position. For instance, grammaticalized topic (GTOP) and grammaticalized focus (GFOC) can occupy the sentence-initial position as well, constituting marked clausal structures via topicalization or pseudo-clefting. This is illustrated in the following examples:

(16) watan ga, m-aniq qulih.
    Watan Top AV-eat fish
    ‘As for Watan, (he) eats fish.’ <E>

(17) watan qu’ m-aniq qulih. 42
    Watan Nom AV-eat fish
    ‘(The one who) eats fish is Watan.’ <E>

Example (16) is a sentence wherein the original subject watan is topicalized and placed in the topic position followed by the topic marker ga. 43 In example (17), the original subject watan is pseudo-clefted and becomes a nominal predicate, preceding the clausal subject maniq qulih ‘eat fish’. Topicalization and pseudo-clefting both lead to the superficial SVO word order. The in-depth discussion on the properties of these two constructions is left to Chapter 7 & 8.

After having a primary understanding of the unmarked word order, the following

---

42 One may suggest translating this sentence as “Watan is the one who eats fish”. In fact, this suggested translation does not match not only the syntactic structure but also the information structure of the original sentence because watan is a nominal predicate and it conveys contrastive information in the original sentence.
43 Egerod (1966, 1980, 1999) distinguishes ga/gaa from ga’. The former is a particle for non-final statement and the latter is a particle for precentral exposure. What is more confusing, it is also mentioned in Egerod (1999) that ga/gaa indicates “the topic of a following comment” and that ga’ functions as a theme particle with the meaning as for and occasionally functions like ga. The present thesis does not make such kind of distinction. All of ga, gaa, and ga’ are analyzed as free variants of ga. There are two reasons. First, the glottal stop in the word final position (specifically, the coda position of the stressed syllable) of some words may not exist at the input level/the underlying representation in phonology. Their existence may just fit into the weight requirement. For example, although both aki’ ‘exist‘ and kita’ ‘see’ both have a word-final glottal stop, only the former keeps this consonant in the affixed form kitan whereas the latter does not in its affixed form kitan. Second, all of the functions of different variants of ga can be systematically accounted for by pragmatic extension. Once a topic marker loses its foregrounding effect, it is not unexpected that its function will be extended to a kind of particle, discourse marker or conjunct for non-final statement followed by other statements.
two sections will delve into how a simple clause is analyzed by means of phrase structure. Two theoretical issues will be touched upon: one is on whether there exists an I’ or an internal S in an AV clause and the other is on whether an oblique actor is an adjunct or a specifier.

### 3.2.2 I’ or Internal S

The rigidity of word order shows that the phrase structure in Squiql Atayal is not as flat as that in non-configurational languages like Warlpiri (Cf. Chapter 2). Take the sentence in (18) below for instance. Its phrase structure cannot be the one in (19) because the flatness of (19) allows the possibility of other word orders, such as SVO, OSO, VSO and so forth. Put differently, there is a hierarchical structure inside this sentence.

(18) m<n>ihiy watan (qu’) tali’.
   AV<Perf>hit Watan Nom Tali
   ‘Tali hit Watan.’ <E>

(19)

![Diagram]

Logically, there are two possible c-structures, as shown in (20) and (21). In the former, the finite verb mnihiy ‘hit’ and the object watan form the constituent I’ first, which
further constitutes the whole clause IP with the subject *qu’tali’*. In contrast, in the latter, the object *watan* and the subject *qu’tali’* form the constituent S first, which further constitutes the whole clause IP with the finite verb *mnihiy* ‘hit’.

(20)  
\[
\begin{array}{c}
\text{IP} \\
\uparrow = \downarrow \\
\text{I'} \\
\uparrow = \downarrow \\
\text{DP} \\
\uparrow = \downarrow \\
\text{I} \\
\uparrow = \downarrow \\
\text{VP} \\
\uparrow = \downarrow \\
\text{m<n>ihiy} \\
\uparrow = \downarrow \\
\text{AV<Perf>hit} \\
\uparrow = \downarrow \\
\text{watan} \\
\uparrow = \downarrow \\
\text{Watan} \\
\end{array}
\]

(21)  
\[
\begin{array}{c}
\text{IP} \\
\uparrow = \downarrow \\
\text{I} \\
\uparrow = \downarrow \\
\text{S} \\
\uparrow = \downarrow \\
\text{VP} \\
\uparrow = \downarrow \\
\text{m<n>ihiy} \\
\uparrow = \downarrow \\
\text{AV<Perf>hit} \\
\uparrow = \downarrow \\
\text{watan} \\
\uparrow = \downarrow \\
\text{Watan} \\
\end{array}
\]

The former is modified from the phrase structure proposed by H. Chang (1997) for Seediq and Kavalan, while the latter is modified from the phrase structure proposed by Kroeger (1993) for Tagalog. These are all Austronesian languages, genetically and

---

44 Note that VP in the above trees immediately dominates DP, seemingly lacking its head V. They violate the endocentric constraint in MGG (i.e. X’ convention) that every phrasal level is projected from a head within itself (Chomsky 1981, Jackendoff 1977). Because of this constraint, MGG requires head movement, one of its three major movements, so as to make a single main verb combine all features of different phrases. It is not problematic at all in LFG, however, since this framework recognizes the existence of non-configurational languages. It is not imperative for phrases to be endocentric.

Another theoretical explanation provided by Bresnan (2001) is as the following: configurational languages still have to obey endocentricity, but the definition is broadened as in (i):

(i) Endocentricity (Bresnan 2001:134):

Every lexical category has an extended head.

(ii) Extended Head (Bresnan 2001:132):

Given a c-structure containing nodes N, C, and c-to f-structure correspondence mapping φ, N is an extended head of C if N is the minimal node in φ(C) that c-commands C without dominating C.

45 In Tagalog, the verbal predicate does not form a constituent but other types of predicates do (Kroeger 1993). Kroeger’s c-structure is modified by Sadler (1997) and Bresnan (2000, 2001) in their studies on Welsh, a configurational internal-subject language, in which subject precedes VP and follows the initial finite verb/auxiliary. These are studies adopting the LFG framework. From the perspective of MGG, however, there are four major ways to analyze verb-initial languages: phrase structure parameter setting, subject-adjunction, VP-fronting and V/I-fronting (Carnie & Guilfoyle 2000, Carnie, Harley & Dooley...
typologically close to Squliq Atayal in question. Thus, both (20) and (21) are both qualified as the c-structures for sentence (18).

It is argued in the present thesis that the c-structure in (20) is appropriate to Squliq Atayal because of the evidence from temporal adverbials. The positions of temporal adverbials are illustrated in the following examples:

(22) **hira’ (ga)**, m<n>ihiy watan qu’ tali’.
    yesterday Top AV<Perf>hit Watan Nom Tali
    ‘Tali hit Watan yesterday.’ <E>

(23) *m<n>ihiy **hira’** watan qu’ tali’.
    AV<Perf>hit yesterday Watan Nom Tali <E>

(24) m<n>ihiy watan **hira’** qu’ tali’.
    AV<Perf>hit Watan yesterday Nom Tali <E>

(25) *m<n>ihiy watan qu’ **hira’** tali’.
    AV<Perf>hit Watan Nom yesterday Tali <E>

(26) m<n>ihiy watan qu’ tali’ **hira’**.
    AV<Perf>hit Watan Nom Tali yesterday <E>

Temporal adverbials, like **hira’ ‘yesterday’**, can be placed in the sentence-initial position as a GTOP like (22), but they cannot intervene between the finite verb **mnihiy ‘hit’** and its object **watan like (23).** In addition, they can immediately precede the

46 The fact presented here contradicts the data in Hsiao (2004), where the intervention of temporal adverbials between the main verb and its object is grammatical. This discrepancy cannot be attributed to dialectal variation because our data are from the same fieldwork site and certainly from the same dialect.
nominative marker *qu’* like (24) but cannot immediately follow it like (25), the latter of which breaks the constituency of the nominative marker and the subject *tali’*. In addition, as shown in (26), temporal adverbials can be placed in the sentence-final position.\(^{47}\)

The positions of temporal adverbials are correlated with c-structure. Not assuming that there is a universal hierarchy of adverbials as proposed in Alexiadou (1997) and Cinque (1999), the present thesis adopts the traditional analysis that temporal adverbials are adjoined to I’, IP or S since adverbial modification is a process of linearly rearranging reference time and event time in accordance with the meaning of temporal adverbials (Hornstein 1990).\(^{48,49}\) Therefore, the distribution of temporal adverbials discussed above helps determine which c-structure, (20) or (21), is valid. Recall that the main verb and its object form a constituent I’ in (20), while the object and the subject form an internal S in (21). In the former c-structure, it is rational that temporal adverbials cannot intervene between main verb and its object since the head I cannot be adjoined by an adjunct, as the annotated c-structure in (27). Under this analysis, the mapping of temporal adverbials from f-structure to c-structure adheres to a language-particular constraint in (28).

\(^{47}\) According to my informants, sentence-final temporal adverbials are kind of afterthought, appended to the whole sentence.

\(^{48}\) Temporal adverbials can be licensed by aspectual inflections/auxiliaries (Andrews 1982a, 1983). There is no distinction between TP and AspP in LFG, so no matter whether temporal adverbials are licensed by tense or aspect, they are adjuncts of I or IP.

\(^{49}\) Following the analysis by Enç (1986, 1987) that temporal adverbials are referential expressions, Alexiadou (1997) proposed that temporal adverbials were V-complements, a position determined by the event-argument structure, further moving to the specifier of TP. This proposal is not only incompatible with the non-transformational approach adopted in the present thesis but also complicates the original rather simple adjunction mechanism, so it is not taken into consideration in my discussion.
(27)  Constraint on Temporal Adverbials in Squiliq Atayal

Given an ADJ has a function TENSE (α), \( \phi^1 \) (ADJ) is adjoined to the right of the node x, forming a recursive x, and \( x \in \{ I', IP \} \); if ADJ = GTOP, \( \phi^1 \) (ADJ) is adjoined to the node y, and \( y \in \{ CP \} \).

Plainly speaking, adjuncts with the tense property in f-structure correspond to adverb phrase (AP) adjoined to the right of I' or IP in c-structure unless they are GTOPs, as demonstrated in (29) below:
Note that the adjunction to the right is not ad hoc because it accords with the general syntactic property in Squiq Atayal that referential expressions are to the right of verbal elements provided that temporal adverbials are referential expressions as advocated by Enç (1986, 1987).

In contrast, the c-structure in (21), where object and subject are immediately dominated by S, has to resort to an ad hoc stipulation that an adjunct cannot be adjoined to the left of S for the unknown reason, as in (30); otherwise, it does not make a straightforward account of the contradiction that adjuncts can be adjoined to S through
intervening between object and subject as in (31).

When there is an auxiliary in a clause, it must precede the main verb. Most important of all, the intervention of temporal adverbials is not permitted between an auxiliary and the main verb, as shown in (33). The grammaticality of other positions is just like the cases of single main verb discussed above. To be more specific, temporal
adverbials can be GTOPs like (32), and they can precede the nominative marker *qu’*,
like (35), or follow subject, like (37). Conversely, they cannot intervene between the
main verb and its object, like (34), or between subject and its nominative marker, like
(36).

(32) **hira’**  ga wal  m-ihiy watan qu’ tali’.
    yesterday Top Aux.Pt AV-hit Watan Nom Tali
    ‘Tali hit Watan yesterday.’ <E>

(33) *wal  **hira’**  m-ihiy watan qu’ tali’.
    Aux.Pt yesterday AV-hit Watan Nom Tali <E>

(34) *wal  m-ihiy **hira’**  watan qu’ tali’.
    Aux.Pt AV-hit yesterday Watan Nom Tali <E>

(35) wal  m-ihiy watan  **hira’**  qu’ tali’.
    Aux.Pt AV-hit Watan yesterday Nom Tali <E>

(36) *wal  m-ihiy watan qu’  **hira’**  tali’.
    Aux.Pt AV-hit Watan Nom yesterday Tali <E>

(37) wal  m-ihiy watan qu’  tali’  **hira’**.
    Aux.Pt AV-hit Watan Nom Tali yesterday <E>

    In the same vein, the c-structure analysis in (21) cannot account for the
impossibility of the adjunction of temporal adverbials to the left of S, as shown in (38),
without any ad hoc stipulations.
In contrast, the c-structure analysis in (20) wins out again because it provides a consistent account of their different positions, as shown in (39), by generalizing them into the simple constraint in (28).
The above evidence shows that the internal structure of an AV clause contains a hierarchy wherein finite verb and its object constitute a constituent without the possibility of being intervened by temporal adverbials. The alternative analysis that subject and object constitute an internal S does not work since it is unable to account for the position of temporal adverbials.

3.2.3 Specifier vs. Adjunct

After clarifying the constituency of AV clauses, this section turns to discuss the
c-structure of NAV clauses, the debatable issue of which is the c-structural status of oblique actors. Take the NAV sentence in (40) for example.

(40) bhy-an na’ tali’ qu’ watan.
    hit-NAV Obl Tali Nom Watan
    ‘Watan was hit by Tali.’ <E>

How can the oblique actor na’ tali’ be analyzed? Because of the restriction of phrase structure, there are two possible theoretical analyses. The first one is the specifier analysis that the oblique actor na’ tali’ is a VP-specifier, as shown in (41). This analysis is modified from previous analyses proposed for other Formosan languages, such as H. Chang (1997) for Seediq and Kavalan, M. Chang (2004) for Tsou, and Mei (1994) for Mayrinax Atayal.50

(41)

The second analysis is the adjunct analysis that the oblique actor na’ tali’ is an I’-adjunct,

---

50 These analyses adopted the framework of MGG, proposing that oblique actors were based-generated at the specifier of VP and grammatical subjects moved to the specifier of IP or even CP for various theoretical reasons and mechanisms. Since LFG allows no vacant V, the empty head of VP is removed.
as shown in (42). This analysis is in accordance with the traditional analysis of European languages that oblique arguments are adjuncts.

(42)

It is claimed in the present thesis that the adjunct analysis is more favorable than the specifier analysis. The evidence comes from the distribution of temporal adverbials. It is well-known that adjuncts are freely-ordered, like the English examples in (43) from Jackendoff (2002:256).

(43)  
   a. Sam struck gold last night in Alaska with his trusty pick.  
   b. Sam struck gold in Alaska last night with his trusty pick.  
   c. Sam struck gold with his trusty pick last night in Alaska.

The three adjuncts in these sentences, last night, in Alaska and with his trusty pick, are commutable in terms of their positions. The same phenomenon occurs in NAV clauses in Squliq Atayal as well. The relative order between oblique actors and temporal adverbials are rather flexible in NAV clauses. Temporal adverbials can precede or follow oblique actors, as exemplified in (44) and (45) respectively.
(44) bhy-an hira’ na’ tali’ qu’ watan.
    hit-NAV yesterday Obl Tali Nom Watan
‘Watan was hit by Tali yesterday’ <E>

(45) bhy-an na’ tali’ hira’ qu’ watan.
    hit-NAV Obl Tali yesterday Nom Watan <E>

This commutability reflects that the oblique actor na’ tali’ is an adjunct in nature. The specifier analysis faces the trouble in explaining this fact without any further stipulation, such as a rather complex licensing condition of temporal adverbials. For instance, the specifier analysis must allow the temporal adverbial hira’ ‘yesterday’ to adjoin to VP in order to account for the commutability between hira’ ‘yesterday’ and the oblique actor na’ tali’, as shown below:

(46)
However, it is hard to justify why temporal adverbials can be VP-adjuncts. Conversely, the adjunct analysis does not have this trouble since the iteration of adjunction exactly accounts for the commutability between *hira’ ‘yesterday’* and the oblique actor *na’tali’, as shown below:

Only when the oblique actor is an I'-adjunct is it straightforward to explain the commutability phenomenon.

The advantage of the adjunct analysis is even more outstanding when it comes to the NAV clauses with auxiliaries, wherein main verb is the complement of auxiliary, as presented in (48).
(48) \( \text{wal bhy-an na’ tali’ qu’ watan.} \)

\begin{verbatim}
Aux.Pt hit-NAV Obl Tali Nom Watan
\end{verbatim}

‘Watan was hit by Tali.’ <E>

The specifier analysis predicts that the temporal adverbial \textit{hira’} ‘yesterday’ can intervene between the auxiliary \textit{wal} and the main verb \textit{bhyan} ‘hit’ on the assumption that temporal adverbials can be adjoined to VP. This is illustrated below:

(49)

In contrast, the adjunct analysis predicts that the temporal adverbial \textit{hira’} ‘yesterday’ cannot intervene between the auxiliary \textit{wal} and the main verb \textit{bhyan} ‘hit’ because temporal adverbials cannot be adjoined to I, as illustrated below:
In fact, the prediction of the adjunct analysis is born out. The temporal adverbial *hira*’ ‘yesterday’ cannot intervene between the auxiliary *wal* and the main verb *bhyan* ‘hit’, as shown in the example (51). The specifier analysis makes a wrong prediction because the restriction on the position of temporal adverbials is looser under this analysis.

(51) *wal  hira’  bhy-an  na’ tali’ qu’ watan.*
    Aux.Pt yesterday hit-NAV Obl Tali Nom Watan <E>

To sum up, the discussion in Section 3.2 shows three important facts as to the c-structure of unmarked clauses. First of all, bivalent clauses in Squiliq Atayal have the unmarked word order VOS, i.e. verb-object-subject in AV clauses and verb-oblique-subject in NAV clauses. The rigidity of word order manifests the hierarchical structure inside a clause. Second, the main verb of an AV clause and its object (patient) are sister nodes in c-structure, forming I’ (or VP when there is no
auxiliary) without the possibility of the intervention of temporal adverbials. The subject and the object of an AV clause are not sister nodes; that is to say, there is no internal S under the whole clause. Third, with regard to NAV clauses, oblique actors are \(\Gamma\)-adjuncts instead of VP-specifiers, which is supported by the evidence from the commutability between oblique actors and temporal adverbials. Compared with the specifier analysis, the adjunct analysis gives a more consistent account of the position of temporal adverbials.

### 3.3 F-Structure & A-Structure

Three basic grammatical systems will be sketched in this section: case markers, pronouns and multiple voices. All of them are associated with f-structure and a-structure. It will be discussed whether Squliq Atayal is morphosyntactic accusative or ergative, an intriguing and constantly discussed issue in Austronesian languages. In addition, how the mapping between thematic functions in a-structure and grammatical functions in f-structure is accommodated in LFG will be shown.

#### 3.3.1 Case Markers & Pronouns

According to L. Huang (1993), L. Huang & Hayung (2011), L. Huang et al. (1998), and Rau (1992), there are six types of case markers in Squliq Atayal: nominative, accusative, oblique, genitive, locative and comitative.\(^{51}\) The case marking system is shown in Table 3-1:

---

\(^{51}\) L. Huang & Hayung (2016) has similar classification, with merely three nuances. First, there is no accusative case marker. Only overt case markers are presented. Second, the oblique case is called the instrumental case. Last, these case markers are also divided into two sets: one is for proper name while the other is for common nouns. However, it is mentioned that the distinction is disappearing.
Although accusative case lacks an overt marker whereas other cases have overt marker, nominative, oblique and genitive case markers are omissible because native speakers can distinguish different syntactic roles easily by means of the rigid word order in Squliq Atayal. These case markers precede the noun phrases or the clauses whose case they mark. The form of oblique case markers is homophonous with that of genitive case markers, and as shown soon, this homophony also occurs to personal pronouns.\(^{52}\) Despite the formal sameness, their grammatical functions are different: oblique case marks actors while genitive case marks possessors. Based on the functional distinction, oblique case markers and genitive case markers are listed separately.

Unlike case markers, personal pronouns are complex in Squliq Atayal. There are two sets of personal pronouns: bound and free (Egerod 1966, L. Huang 1989, 1993, L. Huang & Hayung 2016). Bound pronouns are clitics which follow the first element of the predicate, such as auxiliaries, negatives, or finite verbs.\(^{53}\) On the other hand, free pronouns are just like noun phrases. They can be placed in any position wherein noun phrases appear, and they can be preceded by case markers. The whole system of

---

\(^{52}\) The homophony between genitive forms and oblique ones is prevailing in Austronesian languages, which provides important evidence for the claim in Starosta, Pawley & Reid (1982, 2009[1981]) that from the diachronic perspective, voice markers were derived from nominalization.

\(^{53}\) Holmer (1993) proposes that the 1ˢᵗ and 2ⁿᵈ person bound pronouns is cliticized to C⁰ and the 3ʳᵈ person bound pronouns are fused preposition phrases adjoined to IP. This proposal is unable to account for the following sentence unless it is assumed that the temporal adverbial is licensed by I’ or that the 3ʳᵈ personal bound pronouns are licensed by a layer higher than TP (tense phrase), which licenses the temporal adverbial:

(i) *bhy-an hira’ -nya’ qu’ watan.
hit-LV yesterday 3Obl Nom Watan
Intended for ‘Watan was hit by him yesterday.’ <E>

---

Table 3-1. *Case Markers in Squliq Atayal*

<table>
<thead>
<tr>
<th>Nominative</th>
<th>Accusative</th>
<th>Oblique</th>
<th>Genitive</th>
<th>Locative</th>
<th>Comitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• qu’</td>
<td>• Ø</td>
<td>• na’</td>
<td>• na’</td>
<td>• squ’</td>
<td>• ki’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• nqu’</td>
<td>• nqu’</td>
<td>• sa</td>
<td></td>
</tr>
</tbody>
</table>

---

Nominative: 1ˢᵗ and 2ⁿᵈ person pronouns appear as free pronouns.
Accusative: 1ˢᵗ and 2ⁿᵈ person pronouns appear as bound pronouns.
Oblique: 1ˢᵗ and 2ⁿᵈ person pronouns appear as bound pronouns.
Genitive: 3ʳᵈ person pronouns appear as bound pronouns.
Locative: 3ʳᵈ person pronouns appear as bound pronouns.
Comitative: 3ʳᵈ person pronouns appear as bound pronouns.
personal pronouns is illustrated in Table 3-2.

Table 3-2. Personal Pronouns in Squiliq Atayal

<table>
<thead>
<tr>
<th>Person</th>
<th>Number</th>
<th>Nominative</th>
<th>Oblique</th>
<th>Genitive</th>
<th>Neutral</th>
<th>Locative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Sg.</td>
<td>-saku’/-ku’</td>
<td>-maku’/-mu’/-ku’</td>
<td>-maku’/-mu’/-ku’</td>
<td>kun/kuzi’ng</td>
<td>knan</td>
</tr>
<tr>
<td>2nd</td>
<td>Sg.</td>
<td>-su’</td>
<td>-su’</td>
<td>-su’</td>
<td>isu’</td>
<td>sunan</td>
</tr>
<tr>
<td>3rd</td>
<td>Sg.</td>
<td>----------</td>
<td>-nya’</td>
<td>-nya’</td>
<td>hiya’</td>
<td>(hiyan)</td>
</tr>
<tr>
<td>1st</td>
<td>Pl. Inc</td>
<td>-ta’</td>
<td>-ta’</td>
<td>-ta’</td>
<td>ita’</td>
<td>(itan)</td>
</tr>
<tr>
<td>1st</td>
<td>Pl. Exc</td>
<td>-sami’</td>
<td>-myan</td>
<td>-myan</td>
<td>sami’</td>
<td>(sminan)</td>
</tr>
<tr>
<td>2nd</td>
<td>Pl.</td>
<td>-simu’</td>
<td>-mamu’</td>
<td>-mamu’</td>
<td>simu’</td>
<td>(smunan)</td>
</tr>
<tr>
<td>3rd</td>
<td>Pl.</td>
<td>----------</td>
<td>-nha’</td>
<td>-nha’</td>
<td>lha’</td>
<td>(hgan)</td>
</tr>
</tbody>
</table>

Personal pronouns in Squiliq Atayal reflect the differences in person, number and case. Bound pronouns have three cases, nominative, genitive and oblique, but they do not have accusative. Genitive pronouns and oblique ones are homophones as well. When a verb has two bound pronouns as its arguments, their relative order complies with the following three constraints: (1) the first and second personal pronouns precede the third personal pronouns; (2) for the first and second personal pronouns with the same phonological weightness, nominative pronouns precede oblique pronouns; (3) for the first and second personal pronouns with the different phonological weightness, monosyllabic pronouns precede disyllabic pronouns (Liao 2005). Free pronouns have two cases, neutral and locative, but most of the locative ones are lost in the Jianshi variety of Squiliq Atayal, the target subdialect studied in the present thesis, as bracketed in this table. Personal pronouns are not exemplified here because they are observed in most of examples shown in the present thesis. In addition, Squiliq Atayal is a pro-drop
language in linguistic typology so except for genitive ones they can be omitted.

3.3.2 Multiple Voices

From the perspective of linguistic typology, one of the special characteristics of Squiliq Atayal is its multiple voice system, also a characteristic observed in many Austronesian languages.\(^{54}\) In previous research on Atayal, there is a terminological inconsistency.\(^{55}\) Some works, such as Egerod (1965, 1966, 1980, 1999), L. Huang (1991, 1993, 1995a) and Rau (1992), call these verbal derivations “voices”, while other works, such as L. Huang (1995b, 1996a, 2000b), L. Huang & Hayung (2016), Mei (1994), Zeitoun (2001), call them “foci/focuses”. In order not to arouse confusion, the present thesis adopts “voices” for the verbal derivations agreeing with subjects in their semantic roles and keeps the term “focus” for information-related discussion. This language has four voices: actor voice (AV), patient voice (PV), locative voice (LV) and circumstantial voice (CV).\(^{56}\) The last three voices are grouped as non-actor voice (NAV) or undergoer voice because their subject cannot be an actor. The agreement between voice marking and the semantic roles of subject is illustrated in the following examples from Liu (2004b:27):

\(^{54}\) In addition to Austronesian languages, multiple voice system is also observed in Western Nilotic languages, like Dinka and Kurmuk (Andersen 1991, 2015).Grammatical subjects in these languages are also grammaticalized topics, so the agreement between verbs and their subjects is termed “orientation” in literature. There are subject-orientation, object-orientation and adjunct-orientation, whereby subject and object refer to agent and patient respectively.

\(^{55}\) This terminological inconsistency exists in Austronesian linguistics as a whole, not merely in the study on Atayal or Formosan languages. The term “focus” came from Healey (1958) and Healey (1960), and it was adopted to emphasize the difference between Austronesian languages and European ones. Blust (2002) compiled an appendix of the use of different terms in previous works: 25 sources used “focus” while 28 sources used “voice”.

\(^{56}\) Following S. Huang (2005), Rackowski & Richards (2005), Ross (2006), and Wu (2007), S. Chen (2007) analyzed LV and CV in Atayal as applicatives. Whether LV and CV are applicatives is beyond the scope of the discussion here. Thus, the present thesis just follows the convention in previous studies on Squiliq Atayal. Besides, following L. Huang (1993, 1995a), the present thesis adopts the term “circumstantial voice” to replace instrumental/beneficiary voice. Blust (2002) suggested a typological distinction inside Austronesian languages: four-term languages with at least four morphologically distinguished voices, three-term languages with three voices, two-term languages with a portmanteau infix –in-, and two-term languages without a portmanteau infix –in-. Squiliq Atayal belongs to the first type of Austronesian languages owing to its four-voice system.
As shown in (52), when the verb maniq ‘eat’ is prefixed with the AV marker m-, the subject tali’ is the actor of the whole event. When the patient of the whole event is a subject, like qulih qasa ‘that fish’ in (53), the verb is suffixed with the PV marker -un.

What is more, if the verb is derived with the LV marker -an, like niqan in (54), the subject is the location of the event. In the same vein, when the verb is marked with the CV marker s-, like sqaniq in (55) or (56), the subject can be an instrument in (55) or a beneficiary in (56).

What is presented in the above examples is merely an ideal agreement between voice marking and semantic roles. As a matter of fact, since the number of semantic roles is above four, the mapping between voice derivations and semantic roles cannot be one-to-one; instead, it is always one-to-many, as illustrated in Figure 3-1, which is modified from S. Chen (2007:26) with the addition of AV. S. Chen’s classification of
semantic roles is based on that of L. Huang’s (2001) work on Mayrinax Atayal.

Figure 3-1 *Mapping between Voices and Semantic Roles*

PV and LV are termed after the typical semantic role of subject with which they agree. AV and CV are termed after the general terms covering these roles: actor for agent, executor and experiencer, and circumstantial for beneficiary, instrument, reason and theme.\(^\text{57}\)

\(^{57}\) Following the semantic map connectivity hypothesis of Croft (2003), S. Huang (2005) proposed that instrument and beneficiary in CV clauses were derived from (transported) theme, which explained why some Formosan languages had lost CV after the loss of the mapping between subject and (transported)
Furthermore, voice markers are always portmanteaux, conveying not only multiple voices but also the distinction in polarity, illocutionary force, tense, aspect and mood (TAM). The verb paradigm of Squliq Atayal is sketched in Table 3-3, wherein C stands for consonant and dots for stem. It can be instantiated with *qaniq ‘eat’* in Table 3-4, wherein affixes are highlighted in italics and the stem *qaniq* as well as its variants are in boldface.

Table 3-3 *Verb Paradigm of Squliq Atayal*

<table>
<thead>
<tr>
<th>Polarity</th>
<th>Affirmative</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illocutionary Force</td>
<td>Affirmative</td>
<td>Negative</td>
</tr>
<tr>
<td>TAM</td>
<td>Realis</td>
<td>Irrealis</td>
</tr>
<tr>
<td>Voice</td>
<td>Imperfective</td>
<td>Perfective</td>
</tr>
<tr>
<td>AV</td>
<td>m/-m-/Ø</td>
<td>m-(i)n-</td>
</tr>
<tr>
<td>PV</td>
<td>-un</td>
<td>-(i)n-/un</td>
</tr>
<tr>
<td>LV</td>
<td>-an</td>
<td>-(i)n...-an</td>
</tr>
<tr>
<td>CV</td>
<td>s-</td>
<td>C-</td>
</tr>
</tbody>
</table>

Table 3-4 *Paradigm of qaniq ‘eat’*

<table>
<thead>
<tr>
<th>Polarity</th>
<th>Affirmative</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illocutionary Force</td>
<td>Affirmative</td>
<td>Negative</td>
</tr>
<tr>
<td>TAM</td>
<td>Realis</td>
<td>Irrealis</td>
</tr>
<tr>
<td>Voice</td>
<td>Imperfective</td>
<td>Perfective</td>
</tr>
<tr>
<td>AV</td>
<td><em>maniq</em></td>
<td><em>mmaniq</em></td>
</tr>
<tr>
<td>PV</td>
<td><em>niqun</em></td>
<td><em>qaniq</em></td>
</tr>
<tr>
<td>LV</td>
<td><em>niqun</em></td>
<td><em>qaniq</em></td>
</tr>
<tr>
<td>CV</td>
<td><em>sqaniq</em></td>
<td><em>qqaniq</em></td>
</tr>
</tbody>
</table>
3.3.3 Grammatical Functions

The identification of grammatical functions (also called syntactic roles) is closely related with the moot issue on whether Austronesian languages are accusative, ergative or somewhere in-between (i.e. split-ergative)(Cumming & Wouk 1987). As one of the Austronesian languages, Squiliq Atayal is without exception. There are two analyses in previous literature: one is the traditional analysis and the other is the ergative analysis. The first one is based on most of previous field linguistic studies on Squiliq Atayal, such as Egerod (1966), Rau (1992) and L. Huang (1993), which provided a complete analysis of the case system and the word order in this language. In this analysis, voices in Squiliq Atayal are divided into two groups: AV and NAV. AV clauses have actors as grammatical subjects marked with the nominative case marker qu’ and patients as grammatical objects without any overt case marker, whereas NAV clauses have non-actors as grammatical subjects marked with the nominative case marker qu’ and actors as obliques marked with the oblique case marker na’, as demonstrated in (57). Because of the consistent case-marker of grammatical subject, Squiliq Atayal is viewed as an accusative language.

(57) Traditional Analysis

```
Vi
m-V
(qu’) NP
------------------------
Actor
------------------------
Subject
```

(semantic role)  (syntactic role)

58 As for this issue, it may not be a good idea to make an overall claim—for instance, all Austronesian languages are ergative—by lumping them together with fragmented evidence unevenly collected from several languages since the diachronic development of each language is long enough to derive different syntactic characteristics. The solution may be found either by deeply investigating each language or by designing a checklist of all criteria to analyze whether all Austronesian languages satisfy them or merely part of them.
The ergative analysis is mainly based on Starosta (1988, 1999, 2002, 2009[1997], 2009[1998]), Rau (2000), S. Huang (2005) and S. Huang & Tanangkingsing (2011), which proposed that NAV clauses were canonically transitive clauses while AV clauses were intransitive antipassive clauses. Their frameworks were not the same: Starosta adopted the lexicase dependency grammar while Rau and S. Huang and his associates adopted the function-based approach. Regardless of the difference in theory, their evidence came from case marking, split-subjecthood, transitivity and topicality.\(^{59}\) First, the non-existence of accusative bound pronouns and of overt accusative case marker for common nouns manifests that the patient of AV clauses is not the object of the predicate. Second, the actor of NAV clauses carries some properties belonging to subject, such as the addressee of imperatives and anaphoric prominence. Third, NAV clauses have higher transitivity than AV clauses. Last, the patient of AV clauses has low topicality while the actor of NAV clauses has high topicality. S. Huang (2005) and S. Huang & Tanangkingsing (2011) proposed that actors in NAV clauses, marked with the oblique case marker \(na'\), were ergatives while patients with the case marker \(qu'\) were nominatives. In addition, actors in AV clauses were grammatical subjects, marked with

---

\(^{59}\) Transitivity is a global discourse-pragmatic property of an entire clause which shows to what extent an activity is transferred from an agent to a patient. It consists of ten components: participants, kinesis, aspect, punctuality, volitionality, affirmation, mode, agency, affectedness of O and individuation of O (Hopper & Thompson 1980).
the case marker *qu’*, and patients were extended objects (i.e. obliques) without case-marking. This is illustrated in (58). In this analysis, Squliq Atayal is an ergative language, where NAV clauses are transitive and ergative in their case-marking while AV clauses are intransitive and antipassive in their case-marking.\(^{60}\)

(58) Ergative Analysis

<table>
<thead>
<tr>
<th>Case Marking</th>
<th>(qu’) NP</th>
<th>(qu’) NP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vi</em></td>
<td>Actor</td>
<td>Actor</td>
</tr>
<tr>
<td><em>m-V</em></td>
<td>Subject</td>
<td>Subject</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Marking</th>
<th>NP</th>
<th>(qu’) NP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Antipassive</em></td>
<td>Patient</td>
<td>Actor</td>
</tr>
<tr>
<td><em>(AV)</em> <em>m-V</em></td>
<td>Oblique</td>
<td>Subject</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Marking</th>
<th>(na’) NP</th>
<th>(qu’) NP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Active</em> <em>(NAV)</em></td>
<td>Actor</td>
<td>Non-Actor</td>
</tr>
<tr>
<td><em>V-un</em></td>
<td>Subject</td>
<td>Object</td>
</tr>
<tr>
<td><em>V-an</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>s-V</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both of the above analyses are able to explain why the non-actor of NAV clauses has the morphosyntactic properties of grammatical subject, such as topicalization, relativization, verb agreement, etc., which will be presented later in Chapter 9. For the traditional analysis, it goes without saying that a grammatical subject has these properties. For the ergative analysis, the patient/non-actor of a transitive clause in a

\(^{60}\) Studying Tsou and Seediq, H. Chang (2003) proposed the split-ergative analysis in that AV clauses were accusative while NAV clauses were ergative. This analysis is a compromise between the traditional analysis and the ergative analysis. Although this analysis is valid for Tsou and perhaps for other Formosan languages, as will be discussed in Chapter 8, it is still invalid for Squliq Atayal because of the same reasons presented below.
(syntactically) ergative language has the morphosyntactic properties of grammatical subject, a fact observed in Dyirbal (Dixon 1972). Their divergences mainly lie in the identification of syntactic roles which the patient of AV clauses and the actor and the non-actor of NAV clauses play, as highlighted in the shadowed boxes.

A thorough research on whether Squiliq Atayal is an ergative language is far beyond the scope of the present thesis, but a stance must be taken since this issue and the identification of grammatical functions are mutually dependent. The present thesis favors the traditional analysis over the transitivity one; put differently, it is believed that Squiliq Atayal is a morphosyntactically accusative language. The ergative analysis has the following weaknesses. First of all, both transitivity and topicality are relative discourse-pragmatic properties, so the tendencies they show should not be employed as a criterion for clear-up syntactic categorization, as shown below:

(59) Jerry likes beer. (Hopper & Thompson 1980:253)

(60) Jerry knocked Sam down. (Hopper & Thompson 1980:253)

Syntactically, both of the above sentences in (59) and (60) are transitive without any doubt. However, the latter is much more transitive than the former because the event is a telic and punctual action and the patient Sam is referential, animate, proper and totally affected. The difference in discourse-pragmatic transitivity does not match the syntactic notion of being transitive.

In the same vein, topicality does not work well on the differentiation between syntactically transitive clauses and syntactically intransitive clauses, as exemplified below:

(61) Corruption and nepotism destroy a government.
(62) The European Union is grappling with this moral problem.

Although the object in (61) *a government* is an indefinite noun phrase conveying new information, characterized by low topicality in discourse-pragmatics, it cannot override the fact that the main verb *destroy* is syntactically transitive. Likewise, although the object in (62) *this moral problem* is a definite noun phrase conveying old information, characterized by high transitivity in discourse-pragmatics, it cannot override the fact that the main verb *grapple* is syntactically intransitive.

What both of transitivity and topicality show is merely a tendency of relative degree, and a tendency cannot answer the categorical question on whether a sentence is transitive or not. Take the following non-linguistic down-to-earth fallacy for example. We cannot make a categorical judgment that some centenarian whom we bumped into on the ANU campus is a female merely because of the tendency that females have longer life than males. The misuse of evidence from relative properties to make a clear-cut categorization mistakenly excludes other possibilities such as the case of male centenarians, the case of transitive clauses with low transitivity, or the case of objects with low topicality. Turn to the linguistic case of Squiliq Atayal. It is methodologically misleading that the proponents of the ergative analysis like to apply transitivity and topicality to answer categorical questions on whether or not NAV clauses are syntactically transitive and on whether the patient in NAV clauses are object or oblique. Questions are properly answered only when we know the nature of questions themselves. Otherwise, as stated in Cumming & Wouk (1987), “the notion of discourse ergativity…obscure rather than clarify the functional bases of the syntactic and morphological variation found in these languages [(i.e. Austronesian languages)] (p. 271).”

The second weakness of the ergative analysis is the correlation between voice and
aspect/event type. Passive voice tends to be perfective aspect and stative event (Comrie 1976, Keenan & Dryer 2007). In Squliq Atayal, AV clauses without any auxiliary are imperfective and dynamic while NAV clauses without any auxiliary are perfective/irrealis and stative. Everything being equal, it is more reasonable to associate NAV clauses with passive voice than to associate AV clauses with active voice. It will be difficult to explain the reverse association between voice and aspect/event in Squliq Atayal if it is claimed that NAV is canonically active voice. The dichotomy purely based on discourse criteria always contradicts the general tendencies discovered in morphosyntactic literature.

Third, the ergative analysis ignores the inconsistence of voice marking. Both in accusative languages and in ergative languages, intransitive clauses have the same voice marking as canonically transitive clauses. Comparing the traditional analysis in (57) with the ergative analysis in (58), one can tell that it is the former that really fits into this universal condition on voice marking. That is to say, the ergative analysis has trouble explaining why the voice marking of canonically transitive clauses is different from that of intransitive clauses. It further proves that this language is not a morphosyntactically ergative language. Rather, Squliq Atayal is a morphosyntactically accusative language.

In addition to the above proposal-internal weaknesses of the ergative analysis, the following paragraphs will identify the grammatical functions of the patient in AV clauses, the actor in NAV clauses, and the non-actor in NAV clauses with empirical evidence which further disfavors the ergative analysis. As for those grammatical functions without controversy, the subject of intransitive clauses and the subject of AV clauses, they will be not discussed here.

**Object of AV Clauses**

The main proof that the patient of AV clauses is an object comes from constituency.
As presented earlier in Section 3.2.2, main verb and its patient form a constituent in an AV clause, which cannot be intervened by temporal adverbials. This empirical evidence runs afoul of the ergative analysis that AV clauses are antipassive clauses with extended objects (i.e. obliques). The firm constituency between verb and grammatical object is ubiquitous in configurational languages such as English (Marantz 1984, Tomlin 1986).^61^ The cognitive basis is the chunking of sequential experiences which have high frequency and are more accessible to the memory organization (Bybee 2007, 2010). The motive can be further attributed to the verb’s directly affecting its grammatical object semantically. Without any further evidence, we cannot claim that Squiliq Atayal is an exception to this typological and cognitive tendency. Therefore, the traditional analysis is validated again in terms of constituency.

Nevertheless, the ergative analysis bases its claim that the patient of AV clauses is oblique on two pieces of evidence. First, the patient of AV clauses can be preceded by a locative case marker, *squ’/salte*, which shows that it is an oblique (Huang & Tanangkingsing 2011). This evidence is not attested by the informants of the present thesis because the existence of locative case makers makes a difference in meaning, as illustrated below:

(63) s-m-’ung=saku’ *squ’* qhoniq.
    AV-chop=1SNom Loc tree
    ‘I am chopping (wood) in the forest.’ <E>

(64) s-m-’ung=saku’ qhoniq.
    AV-chop=1SNom tree
    ‘I am chopping trees.’ <E>

^61^ Bresnan (2001) proposed an endocentricity scale based on the typological classification established in Nordlinger (1998). In this scale, the constituency of VP only existed in some configurational languages. In addition, Tomlin (1986) proposed the principle of Verb-Object Bonding, the essential claim of which was that “the object of a transitive verb forms a more cohesive syntactic and semantic whole that does a transitive verb and its subject (p. 4).”
The word *qhoniq* ‘tree’ is preceded by the locative case marker *squ*’ in (63) while it is not in (64). As shown in the English translations, the word *qhoniq* ‘tree’ in the former example is not the patient of the action *sm’unq* ‘chop’ but the place where it is happening. In contrast, the word *qhoniq* ‘tree’ in the latter example represents the real patient undergoing the action. Treating (64) as the case derived from (63) through the omission of the case marker *squ*’ not only misinterpretes the example but also provides misleading evidence for the ergative analysis.

Second, as claimed in the ergative analysis, the null accusative case marker and the non-existence of accusative bound pronouns in the traditional analysis show that there is no accusative case and that the patient of AV clauses is not the object (Cf. Section 3.2.2). The second evidence misleadingly regards case marking as the only morphosyntactic proof of identifying grammatical functions. The non-existence of the accusative case marker does not necessarily represent the loss of the object as grammatical function. If it were true, there would be no differentiation between subject and object in languages without overt case marking, like Mandarin Chinese (Cf. Li & Thompson 1981). Conversely, the non-existence of the overt accusative case marker manifests the firm constituency between a verb and its object because of their cognitive and semantic close relations, as discussed earlier. As for pronouns, the proponents of the ergative analysis only emphasize the non-existence of accusative bound pronouns but ignore the existence of free pronouns as the patient of AV clauses. Bound pronouns are clitics, the morphological forms of which involve lexical idiosyncrasy and cannot be lucidly explained without diachronic evidence.

**Oblique of NAV Clauses**

The actor of NAV clauses is viewed as the subject in the ergative analysis whereas it is viewed as the oblique in the traditional analysis. The present thesis supports the latter because of the following evidence. First, the existence of actorless NAV clauses
proves that the actor of NAV clauses is an oblique since its existence is not obligatory, a property typical of noncore arguments. Squiq Atayal is a language with a null pronoun (also called pro-drop languages), in which it is extremely difficult to attest the non-existence of the actor of NAV clauses in elicitation data since both the inexistent actor and the null actor (pro) have no phonetic form. However, actorless NAV clauses do exist in naturalistic data, as exemplified below:

(65) nanu’ si’ ga s’ang-un, ini’ ltux.
Par if Top shout.down-PV Neg AV.bark
‘Well, if (the dog) is shouted down, (it) will not bark.’ <N>

The above example is a complex sentence composed of the conditional clause si’ ga s’angun ‘if (the dog) is shouted down’ and the main clause ini’ ltux ‘(it) will not bark’. The conditional clause contains a verb suffixed with the PV marker –un. From the discourse-pragmatic context, the patient of the verb clearly refers to the dog, but the actor of the verb does not exist. There is no reason to make a proposal for the existence of a null pronoun. In other words, the actor as the argument of s’angun ‘shout down’ does not exist in this clause, which further shows that its existence is not obligatory. Accordingly, the actor of NAV clauses is a noncore argument—that is, the oblique—as proposed in the traditional analysis.

The second evidence for the actor of NAV clauses being the oblique is from the commutability between the actor and temporal adverbials, a phenomenon already presented in Section 3.2.3. It shows that the actor of NAV clauses is an adjunct. The properties like containing unrecoverable referent in the direct discourse-pragmatic context, being an optional noncore argument in a-structure, and having the word order flexibility of adjuncts in c-structure, all manifest that the actor of NAV clauses is syntactically akin to that of passive clauses in English (Givón 1979a).

The proponents of the ergative analysis may argue that the actor of NAV clauses in
Squiliq Atayal cannot be an oblique since it has some properties which belong to subject in English—to be specific, the addressee of the imperative and the anaphoric prominence. These are split-subjecthood phenomena, and it will be discussed later in Chapter 9 that these two properties are closely related with the fact that actor/agent is the most prominent thematic function, having nothing to do with the grammatical function subject instead.

In addition, the fact that the case marker of the actor of NAV clauses is homophonous with the genitive case marker does not necessarily mean that Squiliq Atayal is a morphosyntactically ergative language (Cf. Starosta 1999, Starosta, Pawley & Reid 2009[1981]). For accusative languages, like French, the actor of their passive clauses may have the same phonetic form as genitive nouns, as illustrated in the following sentence, where the preposition *de* is homophonous with the genitive case marker *de* (Jones 1996, Zhang 2016).

(66) Pierre est aimé **de** tout le monde. (Jones 1996:108)\(^{62}\)

Pierre is loved by all the world
‘Pierre is loved by all the world.’

Homophony may aim to signify the syntactic demotion of the actor in NAV clauses or passive clauses, which is not associated with the classification of ergative/accusative languages.

**Subject of NAV Clauses**

Although the non-actor of NAV clauses has syntactic properties typical of subject, such as topicalization and relativization, which will be discussed later in Chapter 9, the proponents of the ergative analysis may still argue that the object of canonically transitive clauses in ergative languages have these properties as well. However, the ergative

\(^{62}\) The gloss and the English translation are added in the present thesis.
analysis cannot explain why it can trigger agreement but the actor of NAV clauses cannot. As presented in Section 3.3.2, voice marking must agree with the non-actor of NAV clauses. In ergative languages, the agent/actor triggers agreement in active clauses (Falk 2006). If NAV clauses were active and transitive, as proposed by the ergative analysis, the verb should agree with the actor of NAV clauses (i.e. the subject of active clauses in this proposal) instead of the non-actor (i.e. the object of active clauses in this proposal).

So far this section has briefly presented the stance the present thesis takes, that is, the traditional analysis, and the reasons why it is more preferable than the alternative proposal (i.e. the ergative analysis). There is still one thing needing to be clarified. Being a morphosyntactically accusative language only means that this language aligns (transitive) AV clauses with intransitive clauses but does not mean that AV is a default voice like active voice in English while NAV is a derived one like passive voice in English. As argued in Riesberg (2014), Western Austronesian languages are symmetrical voice languages where multiple voices are equally marked. Take Squilq Atayal for example. AV is not unmarked since it is marked with the prefix *m*-, a property as marked as that of NAV. NAV is not derived from AV. If AV were a real default voice, it would not be necessary to have an overt voice marker. Both AV and NAV are equally essential in the grammar of this language. From the viewpoint of theoretical syntax, there is no derivation between AV and NAV, as will be presented in next section.

### 3.3.4 Lexically Determined Mapping

Multiple voice system is theoretically related to the mapping between f-structure and a-structure. Take the AV sentence in (67) for instance. Its f-structure and a-structure are illustrated in (68) and (69) respectively.
In the f-structure in (68), grammatical functions are presented as attributes accompanied with their values: **SUBJ** and **OBJ** stand for subject and object, respectively, and their values are smaller matrixes further composed of two attribute-value pairs, wherein **PRED** stands for predicate and **CASE** stands for case. The value **PRED** of is the lexical sense of the word itself, and the value of **CASE** is the case of the word. The subject of the sentence in (67) is **sayun**, so the value of **PRED** is its sense “Sayun” and the value of **CASE** is its nominative case (Nom). In the same vein, the object **biru’ qasa ‘that book’** has a smaller matrix as its value, in which the sense “that book” is the value of **PRED** and **Acc** is the value of **CASE**. In addition, the **PRED** of the whole f-structure is the sense of the main verb **gmyah ‘open’** and its argument structure because all other grammatical functions are built upon this verb. In addition to the f-structure in, the sentence in (67) has the a-structure in (69), wherein the thematic functions of arguments are presented as attributes as well: the attribute **ACT** stands for actor and **PAT** stands for patient.

One of the NAV counterparts of the sentence in (67) is the PV sentence presented in (70). Its f-structure and a-structure are illustrated in (71) and (72) respectively.
(70) wal gyah-un na’ sayun qu’ biru’ qasa.
   Aux.Pt open-PV Ob1 Sayun Nom book qasa
   ‘That book was opened by Sayun. <E>

(71) f-structure

    SUBJ
    | CASE
    | Nom
    | PRED 'that book'
    | [PRED 'open {SUBJ, OBL}']

(72) a-structure

    SUBJ
    | CASE
    | Obl
    | PRED 'that book'
    | [PRED 'Sayun']

    ACT
    | [PRED 'Sayun']

    PAT
    | [PRED 'that book']

Compared with what is presented about the AV counterpart, there are two major
differences. One of them lies in the f-structure in (71), wherein the PV verb gyahun
‘open’ has an argument structure different from its AV counterpart gmyah ‘open’. The
former requires SUBJ and OBL while the latter requires SUBJ and OBJ. Although the
a-structure in (72) is identical with its AV counterpart in (69), the mapping between
a-structure and f-structure is the other major difference, ACT does not correspond to
SUBJ but to OBL and PAT does not correspond to OBJ but to SUBJ.

The mapping between a-structure and f-structure is handled by the mechanism
called “Lexical Mapping Theory (LMT)”, in which both grammatical functions and
thematic functions are decomposed into two binary distinctive features, [±r] and [±o]
(Bresnan 1994, Bresnan et al. 2016, Bresnan & Kanerva 1989, Bresnan & Moshi 1990,
Bresnan & Zaenen 1990, Levin 1985). The feature [±r] stands for whether or not there
is an inherent limit on what thematic function a grammatical function must correspond
to, and the feature [±o] separates object and secondary object from other grammatical
functions. The values of SUBJ, OBJ, OBL and OBJθ are shown in (73).
Moreover, thematic functions carry intrinsic value of these two features, as shown in (74).

(74) Intrinsic Classification of TFs

Patient-like TFs: \([-r]\)
Secondary patient-like TFs: \([+o]\)
Other TFs: \([-o]\)

With the two binary distinctive features, \([\pm r]\) and \([\pm o]\), a-structure is mapped onto f-structure under the following mechanism:

(75) a-structure to f-structure mapping

SUBJ Mapping 1: A \([-o]\) argument which is the thematically most prominent maps to SUBJ.
SUBJ Mapping 2: \([-r]\) may map to SUBJ.
NonSUBJ Mapping: Other arguments map to the lowest compatible function.

First, the most prominent thematic function which carries the feature \([-o]\), i.e. ACT, is mapped to SUBJ. Once SUBJ mapping is done, other thematic functions will be mapped to the compatible grammatical functions in turn, based on the top-to-down order in (73). As exemplified in (76), after the mapping between ACT and SUBJ (SUBJ mapping 1), PAT maps to OBJ because PAT carries \([-r]\) and so does OBJ (NonSUBJ mapping).
(76) (AV)  
\[ gmyah \text{ ‘open’} : \langle \text{ACT, PAT} \rangle \]

a-structure:  
\[ [\neg o] \quad [\neg r] \]

f-structure:  
\[ \text{SUBJ} \quad \text{OBJ} \]

If SUBJ mapping 1 is inapplicable because of the passivization constraint in (77), then the second subject mapping will be applied to map the thematic function carrying \([\neg r]\) to SUBJ. As exemplified in (78), the symbol Ø indicates that ACT with \([\neg o]\) maps to nothing in the first run, and PAT with \([\neg r]\) maps to SUBJ (SUBJ mapping 2). Finally, the thematic function ACT with \([\neg o]\) maps to the other grammatical function with \([\neg o]\) (NonSUBJ mapping), that is, OBL.

(77) **Passivization Constraint:**

Don’t map the thematically most prominent argument to SUBJ.

(78) (PV)  
\[ gyahun \text{ ‘open’} : \langle \text{ACT, PAT} \rangle \]

a-structure:  
\[ [\neg o] \quad [\neg r] \]

f-structure:  
\[ \text{Ø} \quad \text{SUBJ} \quad \text{OBL} \]

Although LMT works successfully on the voice alternation in asymmetrical languages like English, as argued in Riesberg (2014), it does not work well on that in symmetrical languages, Western Austronesian languages in particular. Her proposal is valid for Squiliq Atayal as well. The mechanism of LMT makes a wrong prediction of the mapping between a-structure and f-structure when it comes to LV, as illustrated in
In (79), the vertical bar in the argument structure separates core arguments ACT and PAT from noncore argument LOC. The passivization constraint in (77) blocks SUBJ mapping 1 so SUBJ mapping 2 links PAT with $[-r]$ with SUBJ, a prediction may not always be born out since it is likely for LV clauses to map LOC to SUBJ. Apparently, there exists a different mechanism in symmetrical languages, as presented in (80), where the mapping between a-structure and f-structure is lexically determined. Put differently, the lexical properties of voice affixes determine what thematic function can be mapped to SUBJ.

This lexically determined mapping nullifies the passivization constraint in (77), directly linking LOC with SUBJ (SUBJ Mapping), as illustrated in (81). The rest of the thematic functions map to compatible grammatical functions: ACT with $[-o]$ maps to OBL and PAT with $[-r]$ maps OBJ.
The proposal of Riesberg (2014) is still problematic because the mapping constraint in (80) makes no account of the fact that \( \text{PAT} \) can map to \( \text{SUBJ} \) in all of PV, LV and CV (Cf. Figure 3-1). To be more specific, why is it that (79) and (81) co-exist in LV clauses? Lexical determination is on the right track but not enough to explain what makes an argument a grammatical subject. In addition, as stated in Riesberg (2014), “one question that naturally arises and that has not yet been explicitly addressed is what determines the choice between the different voices—especially when they are symmetric. This question cannot be answered by LMT on its own, as this module of LFG only provides the technical apparatus for mapping semantics (a-structure) to syntax (f-structure and c-structure)—and vice versa. (p.214)” These two unsolved issues are both related to information structure, as will be discussed later in the present thesis.

### 3.4 Summary

This chapter presents some fundamental empirical morphosyntactic properties of Squiliq Atayal, including its rigid VOS word order, case markers, pronouns and multiple voices. The understanding of these properties contributes a brief but overall picture of the grammar of Squiliq Atayal. In addition, four theoretical issues are touched upon in this chapter. First of all, the evidence from the position of temporal adverbials proves that main verb and its object form a constituent prohibiting the intervention of temporal adverbials and that there is no internal S constituted by postverbal arguments, subject
and object. Second, oblique actors in NAV clauses are analyzed as I'-adjuncts instead of VP-specifiers because of the commutability between oblique actors and temporal adverbials. Third, by comparing the traditional analysis and the ergative analysis, this chapter shows that Squliq Atayal is a morphosyntactically accusative language. The constituency between verb and object and the consistence of voice marking between intransitive clauses and AV clauses are solider evidence than discourse-pragmatic one in the ergative analysis. Last, it is demonstrated that the lexically determined mapping between a-structure and f-structure proposed by Riesberg (2014) is basically on the right track, better than the original version of LMT, but still problematic in precisely selecting one correct thematic function out of many possibilities. This weakness is closely associated with other factors behind the voice alternation and the selection of grammatical subject, especially that of information structure.
4. I-Structure

_I was brought up to believe that the only thing worth doing was to add to the sum of accurate information in the world._

-- Margaret Mead

4.1 Overview

After clarifying the fundamental issues concerning c-structure and f-structure of Squiq Atayal, the present thesis continues to look into the properties of i-structure especially with regard to its primitives and mapping to other representations. In addition to the modified definition of i-structure, it is proposed that the expansion of the feature inventory by adding aboutness and contrast benefits the analysis by identifying more discourse functions than those in LFG literature and that the linearity of discourse functions in i-structure lies in the interplay of violable i-structure constraints. As for the mapping between f-structure and i-structure, the general form of this mapping is one-to-many; that is, one grammatical function corresponds to more than one discourse function. It will be discussed how statistics can be incorporated into each correspondence with the addition of extra symbols and what the two types of the relationship of one-to-many mapping are.

4.2 Primitives

There are three main issues discussed in this section. First of all, the definition of i-structure in LFG is provided. Second, twelve discourse functions in i-structure are identified with the assistance of four discourse features, and the validity of the
feature-based approach will be justified. Last, it is proposed that the arrangement of information in i-structure is the interplay of several violable constraints.

4.2.1 Definition

As reviewed in Chapter 2, one of the most influential definition of information structure is from Lambrecht (1994), as stated below:

INFORMATION STRUCTURE: That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts. (Lambrecht 1994:5)

The above definition is insufficient to be employed for the information structure under the LFG framework. First of all, as a part of theoretical grammar, i-structure in LFG is a “formalized” conceptual representation of information structure, visualized as a set of symbols so as to facilitate theoretical discussion. Second, the information of a sentence is always a fragment of the global information flow, not isolated from the broader discourse-pragmatic context where it exists. This point is missing in Lambrecht’s (1994) definition. Last, pairing information structure with lexicogrammatical structures confines the scope of information structure. Information structure is related with phonology and semantics as well. For example, some African languages make use of tone variation to show whether the information conveyed by some NP is topicalized or not. The i-structure in LFG, as the formal representation of information structure, is definitely not confined in the interactions with lexicogrammar even though they are the focus of the present thesis. By refining Lambrecht’s (1994) definition, the present thesis provides the following definition of i-structure in LFG:
(82) **I-Structure in LFG**

I-structure is the formal representation of discourse-pragmatic information at the sentence level which indicates how it participates in the global information flow. Addressers use it to coordinate with other representations in grammar, in accordance with their mental states in given linguistic and situational contexts, including the spoken/written discourse where the discourse-pragmatic information in question exists.

Under the framework of LFG, information structure is formalized as i-structure composed of attributes and values in matrices (Butt & King 2000, Dalrymple 2001, King 1997, inter alia). The i-structure of one sentence is excerpted from the global information flow in given discourse-pragmatic contexts, as illustrated below:

(83) I-Structure of One Sentence

\[
\begin{bmatrix}
\text{Attribute}_1 & \text{Value}_1 \\
\text{Attribute}_2 & \text{Value}_2 \\
\vdots & \vdots \\
\text{Attribute}_n & \text{Value}_n
\end{bmatrix}
\]

Global Information Flow

\[
\begin{bmatrix}
\vdots \\
\vdots \\
\vdots
\end{bmatrix}
\]

Here, the attributes are symbols representing discourse functions or discourse features, and values can be embedded matrices of i-structure, symbols or semantic/phonetic forms in single quotes.\(^{63}\) Piece by piece, the information from linguistic units is accumulated into information flow.

### 4.2.2 Discourse Functions & Features

After the introduction to the definition of i-structure in LFG, it is necessary to

---

\(^{63}\) The terminology in King (1997), naming the symbols of information-structural notions as “discourse functions”, is followed in the present thesis. Another terminology is “pragmatic functions” in Andrews (2007). In fact, discourse and pragmatics are closely related with each other. The former is chosen just for convenience.
establish the basic units of i-structure, discourse functions and features. As presented earlier in the literature review part of Chapter 2, LFG linguists adopt the feature-based approach to differentiate discourse functions, such as Choi’s (1997, 1999, 2001) two discourse partitioning features [±New] and [±Prom] for topic, tail, contrastive focus and completive focus. These two discourse features cannot meet the need of the present thesis; therefore, in order to cover the phenomena in Squliq Atayal, the inventory of discourse features is expanded by adding another two features, aboutness and contrast, and by converting newness into givenness. The expanded inventory of discourse features is presented below:

(84) Discourse Features

- **Givenness (GIV)**: whether information is presumably possessed by the audience (+) or not (−) in current information flow;
- **Aboutness (ABT)**: whether information is important (+) or unimportant (−) in subsequent information flow;
- **Prominence (PROM)**: whether information is salient (+) or nonprominent (−);
- **Contrast (CONT)**: whether the salience of information is to highlight its referent against others in a set of the same proposition (+) or not (−).

Aboutness (ABT) indicates whether the information conveyed by an entity is important in information flow and relevant to the communicative goal.64 This discourse feature is independent from givenness, as manifested in Tomlin and Rhodes’s (1992) discovery that in a sentence of Ojibwa thematic information (i.e. the knowledge which the addressee assumes to be relevant to the goal of the communicative event) tends to be placed in a position different from shared information (i.e. the knowledge which the addressee assumes to share with the addressee). However, both importance and relevance are abstract notions and how to judge them is rather subjective. The present

---

64 Reinhart’s (1981) and Strawson’s (1964) definition of topics also made use of the term aboutness from the perspective of semantics, in which sentence topic determines the truth value of a sentence. In line with Van Dijk’s (1977) distinction between sentence topic and discourse topic, Reinhart’s (1981) differentiated the aboutness of sentence topic and that of discourse topic. The aboutness in the present thesis refers to the latter.
thesis applies Givón’s measurement of topic persistence to implement the analysis of “aboutness”. In accordance with this measurement, the word subsequent is added to the definition.

The other discourse feature added here, contrast (CONT), marks whether an entity in a context set conveys information which stands out against other set members. As mentioned in Chapter 2, its existence has been recognized by linguists and discovered in many languages. It is claimed in the present thesis that contrast is a discourse feature subordinate to prominence because contrastive information always goes with prominent morphosyntactic structures or prosodic structures cross-linguistically (Givón 1990).

As for the conversion of newness into givenness, it is due to the language-particular fact that Squliq Atayal is more sensitive to givenness than newness, the details of which will be demonstrated later. In fact, newness “allows further breakdown into whether the information is discourse-new or addressee-new (Los, López-Cousou & Meurman-Solin 2012).” Likewise, the broad sense of givenness can be further divided into two categories: (the narrow sense of) givenness (GIV) and oldness, the former of which stands for the information known or presumed to be in possession of an audience whereas the latter of which stands for the information already mentioned in previous discourse, opposite to discourse-newness, as proposed in Erteschik-Shir (2007) and Ward, Birner & Huddleston (2002).65 It is always the case that the information mentioned earlier in discourse is presumed to be in possession of an audience. That is, oldness is always subsumed by givenness. Therefore, this categorization is not adopted in the present thesis.

With the interplay of the above four partitioning discourse features, twelve discourse functions are identified and defined, as shown in Table 4-1: contrastive topic

---

65 The distinction between givenness and oldness is manifested in the discourse-pragmatic difference between the topic marker wa in Japanese and the topic marker nun in Korean. They have been viewed as the equivalent of one another, but the former refers to given information while the latter encodes old information (Lee & Shimojo 2016).
(CSTOP), switch topic (SWTOP), contrastive theme (CSTHM), scene-setting topic (SSTOP), continuing topic (CTTOP), background information (BKINF), contrastive focus (CSFOC), introductory focus (ITFOC), interrogative focus (QFOC), highlighted rheme (HLRHM), presentational focus (PSFOC) and completive information (CPINF). Gray columns mark the inapplicability of contrast under [−PROM].
<table>
<thead>
<tr>
<th>Discourse Functions Proposed in the Present Thesis</th>
<th>[GIV]</th>
<th>[ABT]</th>
<th>[PROM]</th>
<th>[CONT]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrastive Topic (CSTOP)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Switch Topic (SWTOP)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Contrastive Theme (CSTHM)</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Scene-setting Topic (SSTOP)</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Continuing Topic (CTTOP)</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Background Information (BKINF)</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Contrastive Focus (CSFOC)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Introductory Focus (ITFOC)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Interrogative Focus (QFOC)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Highlighted Rheme (HLRHM)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Presentational Focus (PSFOC)</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Completive Information (CPINF)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
</tbody>
</table>
Although some of them look similar to the terms used in some previous linguistic studies, their definitions may not be the same. For instance, highlighted rheme in the present thesis, which represents non-contrastive new information unimportant to the subsequent discourse, is definitely unrelated with “rheme” in Systematic Functional Grammar (SFG), which represents the part of clause excluding the departure point of discourse.

In Table 4-1, Choi’s (1997, 1999, 2001) topic is divided into contrastive topic (CSTOP), switch topic (SWTOP), contrastive theme (CSTHM) and scene-setting topic (SSTOP). Both CSTOP and SWTOP convey given, significant and prominent information but their difference lies in contrast. CSTOP highlights the uniqueness of the referent but usually excludes other referents while SWTOP introduces something not present in the current information flow so as to shift the addressee’s attention, as exemplified in (85) and (86) respectively.

(85)  B:  Tell me about your brother John and Bill. (Erteschik-Shir 2007:48)  
     A:  JOHN is the smart one.

(86)  Once upon a time there was a king with two sons. The older son expected to take over the kingship…As for the younger, he concentrated on studying philosophy at the University. (s. m. Andrews 2007:149)

The reply in (85) puts a stress on John, the information of which is shared between the addressee and the addressee through the dialogue, to express its prominence coming from the contrast between John and Bill. Under the same proposition be the smart one, John is the guy fitting into it while Bill is not. The example in (86) also contains two members. The description of the older son comes first, and then it shifts the discourse line from the older son to the younger son, another given information, by employing the marked syntactic structure left-dislocation. Unlike (85), there is no contrast between the
older son and the younger one since these two expressions are not placed in the same proposition to make a contrast: taking over the kingship and studying philosophy at the University are two different propositions and they cannot be compared. Of course, the two sons can be compared and contrasted on the basis of their attitude, personality, appearances, and so forth, but if the younger is analyzed as “contrastive” information, it will over-broaden the definition of contrast since it is always likely for any two entities in a piece of discourse to be compared and contrasted in some aspects.

As for CSTHM and SSTOP, they both convey unimportant prominent given information. They differ in that the former is contrastive while the latter is not. As exemplified in (87), the preposition phrase in these years is highlighted to contrast with other period of time, but this given information is not repeated anymore in subsequent discourse. Thus, it corresponds to CSTHM in i-structure. Scene-setting topic (SSTOP), also called stage topic, provides tempo-spatial setting for an event, such as the temporal adverbial over the time in (86). Erteschik-Shir (2007) proposed that every clause had its own SSTOP even though it might not be overt, as illustrated in (89) which is a thetic-judgment sentence having no event participants as topics but a covert SSTOP referring to the here-and-now context where this sentence is addressed.

(87) It is in these years that he learned most of the fundamental knowledge of linguistics. After graduating from NTHU, he started his practical training at a high school.

(88) Over the time, the population of these ancient cultures grew several times bigger than it had been before.

(89) It’s snowing. (Erteschik-Shir 2007:17)

The nonprominent given information in Choi (1997, 1999, 2001), i.e. tail, is divided into continuing topic (CTTOP) and background information (BKinF), with a
distinction existing in aboutness: CTTOP is significant to information flow while BKINF is not. This difference is illustrated in (90), in which the pronoun she is a piece of important given information along which the story line extends whereas the pronoun it is a piece of unimportant given information not repeated in subsequent discourse anymore.

(90) …Emma dashed into her study. There was a pen on the desk. She couldn’t wait to write a letter to her husband with it. All her sorrows and joys immediately became appealing words on her letter. Then, she came up with an idea…

Likewise, the existence of aboutness and contrast splits contrastive focus in Choi’s works into four categories: prominent new information is composed of contrastive focus (CSFOC), introductory focus (ITFOC), interrogative focus (QFOC) and highlighted rheme (HLRHM). They are exemplified below:

(91) I went through the whole family—mom, dad, my three brothers, two sisters, the lot—you’d have thought someone would offer to help, but heck, they all refused…It was my great aunt who finally stepped in and…(s. m. Givón 1990:705)

(92) The volume of engine sound became louder and louder. Motorcycle police, a whole battalion (or whatever unit they come in) neared--took over the road--there must have been twenty of them. Behind them there appeared police vans and police buses, one, two, four, six, eight of each. And then, at last, behind these, the American military vehicles began to appear. (Birner & Ward 1998:107)

(93) What is the code opening the door to the genesis of the universe? This puzzle has been fascinating scientists for thousands of years…

(94) After they had travelled on for weeks and weeks past more bays and headlands and rivers and villages than Shasta could remember, there came a moonlit night when they started their journey at evening, having slept during the day. They
had left the downs behind them and were crossing a wide plain with a forest about half a mile away on their left. (Birner & Ward 1998:107)

As shown in (91), the CSFOC conveyed by my great aunt is not only contrastive in that the abovementioned referents did not step in to help with the addressee’s trouble but also significant in that it changes the direction of story line from miserable refusals to a more positive event discussed later. In contrast, the noun phrase police vans and police buses introduced by the there-sentence in (92) is not contrastive but still important to the subsequent discourse, so it corresponds to ITFOC in i-structure. The interrogative word what in (93), a QFOC, has no referent at all, only indicating an information gap between the addresser and the addressee. Usually an identical information gap is not repeated in subsequent discourse unless the addressee avoids giving a reply. The example in (94) makes use of a there-sentence to pinpoint a moonlit night but this new information does not recur in the subsequent discourse, only functioning as the general background for a series of events and scenes. Prominent new information without importance to subsequent discourse is HLRHM in i-structure.

In addition, presentational focus (PSFOC) is differentiated from completive information (CPINF). They were lumped together in Choi’s works. Although they are both new information highlighted neither phonologically nor morphosyntactically, the referent of the former steps onto the stage of information flow and tends to reoccur later, like a man in (95) whereas that of the latter is merely an accessory of information flow, like a kitchen in (96).

(95)  A man goes in the pub. There’s a bear sitting in the corner. He goes up to the, he goes up to the bartender. He says, why is there a bear sitting over there? (Biber et al. 1999:943)

       B: He is eating pizza in the kitchen.
Discourse features are primitives of i-structure, formalized as attributes with binary partitioning values, either positive value (+) or negative one (−), but not both at the same time. Discourse functions are always complexes of more than one discourse feature. The advantage of the feature-based approach lies in their independence and superposability. It helps formalize the fact that all information in information flow may not be best analyzed in a single stream but as in parallel and mutually interactive streams, as exemplified below in (97).

(97) Q: What did John write?  
A: John wrote a poem

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[presupposition]</th>
<th>[focus]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td>−</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[ABT]</th>
<th>[topic]</th>
<th>[comment]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td>−</td>
</tr>
</tbody>
</table>

(focus-presupposition articulation)

(topic-comment articulation)

In the above question-answer pair, the subject-verb part of the answer John wrote has been mentioned in the question so it conveys given information, but the object a poem conveys new information not mentioned in the question. Therefore, from the viewpoint of givenness, the sentence in the answer can be divided into two parts, presupposition (the former, marked in red) and focus (the latter, marked in green). However, with regard to aboutness, the whole question-answer pair is mainly about the entity John, which is important in this interaction. Thus, there is another way to partition the same sentence into two parts, topic (the former, marked in blue) and comment (the latter, marked in orange). The focus-presupposition articulation and the topic-comment articulation are superposed on the same sentence. In the Prague School, this sentence
can be viewed to consist of three parts, i.e. theme-transition-rhyme (Firbas 1974). The feature-based approach successfully achieves the same tripartition without stipulating this tripartition as the only universal organization of information structure. Similarly, if both grammatical subject and object convey given and significant information, functioning as topics—usually subject as the primary topic and object as the secondary one, as discussed in Dalrymple & Nikolaeva (2011)—, then the traditional topic-comment bipartition will face the same challenge that it is uncertain which part the topical object belongs to. Most of the time the values of these attributes (discourse functions/features) are determined by the discourse-pragmatic contexts where the piece of information conveying them exists unless they are specified as lexical properties or as mapping constraints in other representations. In addition, as pointed out by Choi (2001:92), “[o]ne of the advantages of this feature-based information structure is that it can crossrefer to more than one distinct information type.” This advantage will be manifested in the mapping theory proposed in Chapter 8.

It is possible to propose more features, such as expectedness and exhaustiveness, in order to constitute more discourse functions.66 Because of their inapplicability to Squliq Atayal, they are not elaborated in the present thesis despite their significance to other languages. Basically, the present thesis only focuses on givenness, aboutness, prominence and contrast.

66 Jin (Chinese language) has different grammaticalized topic markers for expected topics and switch topics (Guo 2008). The topicality of expected topics is predictable from the immediately preceding discourse, while that of switch topics is not since the entity a switch topic introduces is not referred to recently even though it has been introduced before (Andrews 2007). Interestingly, DeLancey (1997, 2001) advanced “mirative” as a grammatical category independent from evidential, which marks the non-expected and surprising information conveyed by a proposition as a whole. The existence of mirative seems to provide another empirical evidence for the discourse feature EXP. Nevertheless, mirative marking is always related with other grammatical/pragmatic properties, so it is still controversial whether mirative can be viewed as an independent grammatical category (Cf. Aikhenvald 2012, Hengeveld & Olbertz 2012, Hill 2012, Lazard 1999). Perhaps the real discourse features carried by mirative marking vary from language to language, and EXP is merely one of them. In languages like English and Hungarian, it is possible that contrast is divided into two features, exhaustiveness and exclusiveness (É. Kiss 2002, Horn 1981, secondhand cited from Lambrecht (1994)). To my interpretation, exhaustiveness and exclusiveness equal contrastiveness and restrictiveness respectively in Erteschik-Shir (2007). These two features are not adopted here partly because they have no bearing on Squliq Atayal, and partly because it is still controversial whether this distinction really matters.
4.2.3 I-Structure Constraints

It has been proposed in linguistic literature that the arrangement of information follows some principles, the most famous one of which is the given-new principle that known information precedes new one (Chafe 1974, Givón 1984, among others). In addition, Givón (1984) and Mithun (1987) proposed the newsworthiness discourse principle that newsworthy and unpredictable information precedes news-worthless and predictable information. These two principles sound contradictory to each other, but usually they co-occur in one language and even have effect on the same position. For instance, the clause-initial position in Old English could see the effect of these two principles, with the occurrence of either given or unpredictable information, while the effect of given-new discourse principle gradually faded out from the clause-initial position in the history of the English language (Los 2012).

If discourse principles were inviolable, they could not co-exist with each other for a long while and the competition between them would not fade out on a gradual basis. Hence, in accordance with the discourse principles in linguistic literature, the present thesis proposes the following violable i-structure constraints:

(98) **Given-First**

a. SCHEMA : *[[GIV −] [GIV +]]

b. DEFINITION : When two pieces of information are adjacent, given information must come first if there is some.

---

67 What Los (2012) investigates is the clause-initial marked position, i.e. the left peripheries in MGG. According to Ge (2008), the first element in a clause, that is, “theme” in SFG, became less various and accommodated more grammatical subjects throughout the history of the English language; on the other hand, the amount of thematic transformations, such as passivization, clefting and pseudo-clefting, increased. These two studies dovetail each other, showing that the alternation of word order as a syntactic strategy to represent different discourse functions gradually decreased from Old English to Present-Day English.
(99) **Aboutness-Last**  
   a. SCHEMA : *[[ABT +] [ABT −]]  
   b. DEFINITION : When two pieces of information are adjacent, important information to subsequent information flow must come last if there is some.

(100) **Prominence-First**  
   a. SCHEMA : *[[PROM −] [PROM +]]  
   b. DEFINITION : When two pieces of information are adjacent, prominent information must come first if there is some.

The Given-First Constraint in (98) comes from the given-new discourse principle, requiring that given information should be uttered as early as possible, at least preceding new information, so as to connect the current proposition to previous discourse. In contrast, the Aboutness-Last Constraint in (99) aims to postpone the utterance of the information which is important to the subsequent discourse, at least later than unimportant information to the subsequent discourse, so that the current proposition is able to get connected with the subsequent discourse. The Prominence-First Constraint in (100), which originates from the newsworthiness discourse principle, requires that prominent information, usually the information the addressee believes to be newsworthy, unpredictable or event contrastive, should occur in an utterance as early as possible, at least prior to nonprominent information. The schemata of these constraints are presented with the (left-to-right) linear order of discourse features in i-structure. As shown in the schemata, these constraints only reject the worst cases instead of requiring the best cases. Take the Given-First Constraint for example. It does not require that any pair of adjacent information should be given-new, i.e. *[[GIV +] [GIV −]]*. In such case, not only new-given adjacency but also given-given and new-new adjacencies are ruled out. It only rejects the worst case, that is, the new-given adjacency, instead. So do the other constraints.

Although i-structure constraints are like the constraints in OT, all violable in nature,
their ranking is different from the ranking in OT in that even ranking itself is still violable, owing to the difference between mapping type and token. To be more specific, the general mapping between i-structure and c-structure, formed by mapping type, complies with a certain ranking of i-structure constraints, while the reality constituted by mapping token displays the ongoing competition between constraints, which further leads to a violable ranking and even permanent re-ranking. Both i-structure constraints and rankings are usage-based theoretical mechanisms. This will be discussed in the later chapters.

4.3 General Mapping

This section introduces two phenomena observed in the general mapping between two representations, particularly f-structure and i-structure: one is the one-to-many mapping and the other is the type-token relations within it.

4.3.1 One-to-Many Mapping

Squilq Atayal is a configurational language, a fact recapitulated for several times in the present thesis, so when it comes to the correspondence between syntactic modules, it is self-evident that the mapping between f-structure and c-structure is more consistent and straightforward. The mapping from c-structure to f-structure employs the correspondence function $\phi$ while the mapping from f-structure to c-structure employs its inverse correspondence relation $\phi^{-1}$. On the other hand, the correspondences between f-/c-structure and i-structure are always one-to-many especially in terms of the general mapping between them since the interactions between syntax and discourse-pragmatics.

---

68 In LFG theoretical representation, the mapping from each node in c-structure to f-structure is strictly one-to-one: that is to say, one input (a node in c-structure) can only give birth to one output (a grammatical function in f-structure). Therefore, this kind of relation is a function from the viewpoint of discrete mathematics (Dalrymple 2001, Partee, Ter Meulen & Wall 1990). In contrast, the opposite direction of mapping, the one from one grammatical function in f-structure to c-structure, is not a function in discrete mathematics since one grammatical function can correspond to more than one node in c-structure.
are inconsistent in configurational languages. Put differently, a grammatical function in f-structure or a node in c-structure may correspond to two or even more discourse functions in i-structure when the general mapping is taken into consideration. However, when it comes to each individual case, a grammatical function in f-structure or a node in c-structure still corresponds to one discourse function in i-structure.

In order to differentiate individual correspondences in the general mapping between f-/c-structure and i-structure, the annotation of each correspondence is labelled additionally with its proportion.\(^{69}\) The most accurate way is to mark their precise percentage. This way of annotation however has two disadvantages. On the one hand, the percentage presented varies with different statistical studies even though the tendencies it shows may be consistent. As long as the results of different studies are consistent with one another, it does not matter whether their specific numbers are identical. On the other hand, the annotation with percentage deflects readers’ attention from the tendencies they should look into. For the sake of clarity and conciseness, the following symbols of proportion are additionally labelled in front of the LFG annotations.

---

\(^{69}\) Labov (1969) makes a proposal that transformations should be accompanied by variables marking the frequency with which they will be applied under various conditions. Although his proposal is inspiring, the interaction of transformations diminishes the strength of labelling percentage, covering the general tendency presented by the frequency.
In Table 4-2, the symbols of proportion are divided into 8 subtypes: full, dominant, absolutely major, major, minor, absolutely minor, rare, and null. This division is mainly based on the percentage. The full subtype represents the situation where all input tokens get the same output; in other words, there are no exceptions. In contrast, when there is no mapping from an input to a certain output in another representation, the correspondence will belong to the null subtype. These two subtypes are not given any special symbols because it is easily perceived through the number of the correspondence lines marked in mapping whether mapping is one-to-many or one-to-one or even does not exist. Special symbols are given to the other subtypes in one-to-many mapping. The dominant subtype has the percentage close to the full one, between 90% and 100% (90% < & <100%). It is marked with the symbol “☼”. The rare subtype has the percentage close to the null one, between 0% and 10% (0% < & ≤ 10%), and it is marked with the symbol “●”. The rest of the subtypes are in-between the dominant subtype and the rare one, and each of them has a span of approximately 20%. The symbols “♠”, “♣”, “♦”, and “♥” are used to mark absolutely major proportion (70% < &
≤ 90%), major proportion (50% < & ≤ 70%), minor proportion (30% < & ≤ 50%), and absolutely minor proportion (10% < & ≤ 30%), respectively. The division of these subtypes is for the sake of convenience, flexible for the modification of future research.

By labelling these symbols of proportion, the present thesis is able to incorporate the non-discreteness of statistical evidence into LFG theoretical representations.

4.3.2 Type-Token Relations

Within one-to-many mapping, there are two sorts of type-token relations. One is the prototypical type and the other is the independent type. If they are similized with the phonological terminology, the former is like one phoneme with several allophones while the latter is like the co-existence of several phonemes. Phonemes are types and their variants, allophones, are specific tokens. That is to say, there is only one type with several tokens in the prototypical type of one-to-many mapping, while the independent type of one-to-many mapping is merely a collection of several mutually independent types.

The first type of one-to-many mapping is illustrated below:

(101) The Prototypical Type of One-to-Many Mapping

In (101), for the grammatical function GF, it looks like mapping to three discourse functions, DF1, DF2 and DF3, but DF1 and DF 3 are merely the token variations of DF2. Thus, among all the correspondences in the one-to-many mapping, only one
particular correspondence is called “default mapping”, i.e. the mapping between GF and DF2 in (101), since it represents the prototypical mapping relation between two representations. The mapping of grammatical subjects in Squliq Atayal belongs to this case; as will be shown in Chapter 6, grammatical subjects correspond to unmarked continuing topics in i-structure by default, and the one-to-many mapping forms a hierarchy in which the most prototypical correspondence (i.e. the default mapping) occupies the majority and the less prototypical ones occupy lower percentages gradually descending with their extent of deviance. Prototype is the core of this one-to-many mapping. From the perspective of cognitive linguistics, the prototypical type of mapping is one mapping type with token variations. Traugott & Trousdale (2013) differentiates the strong association between an instance and a more general category in a network from the weak one by using a unbroken line to represent the former and a broken line to represent the latter. The same notion can be applied to the present thesis. That is, in the prototypical type of one-to-many mapping, the default mapping relation is a much stronger association between two representations than other non-default mapping relations so the default mapping relation is indicated by the unbroken line while other non-default ones are indicated by the broken line.

The second type of one-to-many mapping is illustrated below:

(102)  **The Independent Type of One-to-Many Mapping**

![Diagram](image)

In contrast to (101), the one-to-many mapping in (102) belongs to the independent type
since the three discourse functions, DF1, DF2 and DF3, are distinct categories, without anyone of them being a token variation. Therefore, all the correspondences between GF and these three discourse functions are mutually independent. Most of the one-to-many mapping shown in Squiliq Atayal belongs to this type.

In reality, the distinction between the above two types of one-to-many mapping is not always clear-cut. With the constant influence of discourse-pragmatic driving forces, the token number of non-prototypical cases may increase to the extent that the default mapping relation is weakened, being not prototypical anymore, with the emergence of two independent categories. For instance, the relative clause in a cleft construction usually conveys given information to the addressee and the clefted nominal new information, as exemplified in (103), where the event that I lost something is given information shared between the addressee but the addressee does not know that that thing is my key.70

(103) It’s my key that I lost. (Lambrecht 1994:70)

Nevertheless, the information types expressed in naturalistic data may be reversed: the relative clause conveys new information while the clefted nominal given information (Lambrecht 1994, Prince 1978), as exemplified in (104), where George Orwell is known to the audience in a lecture while the relative clause definitely contains information new to the audience. The emergence of exceptions or non-prototypical tokens is attributed to the discourse-pragmatic driving force that the addressee makes use of the already existing form-function pair to accommodate more discourse functions.

70 Collins (1991) classifies cleft constructions into three information-structural types: Type 1 contains the clefted constituent with stressed focus and the embedded clause with background; Type 2 contains the clefted constituent with background and the embedded clause with focus; Type 3 contains all-focus information in both of the clefted constituent and the embedded clause. Despite different terminology, the example in (103) belongs to the first type whereas the example in (104) belongs to the second type.
It was George Orwell who said that the best books are those which tell you what you already know. (Lambrecht 1994:71)

Reversely, the constant influence of discourse-pragmatics may augment the token number of prototypical cases until the default mapping relation is established. This direction of development will be discussed in Chapter 9.

### 4.4 Summary

The first part of this chapter discusses how information is formalized as the representation of i-structure in LFG, particularly concerning the proposal of the definition of i-structure, the expansion of the inventory of discourse features and the proposal of i-structure constraints. It is shown that the application of the four binary partitioning discourse features, givenness (GIV), aboutness (ABT), prominence (PROM) and contrast (CONT), constitutes twelve discourse functions: contrastive topic (CSTOP), switch topic (SWTOP), contrastive theme (CSTMH), scene-setting topic (SSTOP), continuing topic (CTTOP), background information (BKINF), contrastive focus (CSFOC), introductory focus (ITFOC), interrogative focus (QFOC), highlighted rheme (HLRHM), presentational focus (PSFOC) and completive information (CPINF). In addition, it is proposed that the representation of i-structure is restricted by the interplay of violable constraints, including the Given-First Constraint, the Aboutness-Last Constraint and the Prominence-First Constraint.

The second part of this chapter is about general mapping, which is the set of individual mappings between f-/c-structure and i-structure. It is displayed in the form of one-to-many mappings. With the labelling of the symbols of proportion on one-to-many mapping, LFG is enabled to incorporate quantitative evidence into the formal representation of i-structure and to present the proportion of each correspondence. In addition, depending upon whether the discourse functions belong to a single category
with a prototypical member or they belong to distinct categories with each of them being independent, one-to-many mapping has two types, the prototypical type and the independent type. Only the former has a default correspondence between f-/c-structure and i-structure while the latter does not.
5. Discourse Evidence: Topicality & Information Chaining

The length of the discourse indicates the distance betwixt the speaker and the hearer. If they were at a perfect understanding in any part, no words would be necessary thereon. --Ralph W. Emerson

5.1 Overview

This chapter tentatively puts aside theoretical discussion on i-structure, focusing on presenting discourse evidence from topicality and information chaining. The study on topicality employs the quantitative method of Givón's works, delving into referential distance (RD) and topic persistence (TP) of nominal expressions in discourse. By adopting inferential statistics, it is tested whether the interplay of syntactic roles and voice alternation has any statistically significant effect on topicality of nominal expressions. As for information chaining, it concerns how the information of two adjacent sentences is chained up through their respective nominal expressions of the same or different syntactic roles. The above two pieces of discourse evidence, despite the application of different research methods, attests the fact that both syntactic roles and voice alternation have a strong association with information structure.

5.2 Topicality

This section starts with the description of data and the statistical research design.
Then it will present the results of the quantitative analysis of topicality based on the distinction between RD and TP. If necessary, examples will be offered as supplements for further discussion.

5.2.1 Data

Of the naturalistic data collected for the present thesis, only 13 narratives are analyzed in the present statistical analysis because these texts do not contain code-switching at the sentence level, which may cause problems about how to count RD and TP in terms of clause number. They altogether contain 829 finite clauses. The frequency of their voice patterns is shown in Table 5-1. There is no copular verb in Squiliq Atayal, so NPred in this table indicates the clause made of nominal predicate.

Table 5-1 Frequency of Voice Patterns

<table>
<thead>
<tr>
<th></th>
<th>NPred</th>
<th>AV</th>
<th>PV</th>
<th>LV</th>
<th>CV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>65</td>
<td>492</td>
<td>113</td>
<td>144</td>
<td>15</td>
<td>829</td>
</tr>
<tr>
<td>Percent</td>
<td>7.8</td>
<td>59.3</td>
<td>13.6</td>
<td>17.4</td>
<td>1.8</td>
<td>100</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td>272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td></td>
<td>32.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

($\chi^2$=859.896, df = 4, p <.00)

AV constitutes the majority of finite clauses (59.3%), significantly outnumbering other categories in statistics (p <.00), whereas NAV altogether amounts to 32.8 percent of the whole. Among NAV, CV is in the decided minority as pointed out in S. Huang (2005) and S. Huang & Tanangkingsing (2011). Nevertheless, whether PV or LV takes first place in NAV depends upon the corpora. PV takes first place in S. Huang (2005), the data of which comes from Formosan Language Archive, Academia Sinica, but LV does
this in S. Huang & Tanangkingsing (2011), which has additional data from the NTU corpus of Formosan languages. The numbers in Table 5-1 show LV clauses are more than PV clauses, a distribution closer to that in S. Huang & Tanangkingsing (2011). In fact, the subject of LV clauses does not necessarily carry the semantic role location; instead, the majority of them are theme or patient. Location as a subject is still in the minority of LV clauses, as demonstrated in Figure 5-1:

Figure 5-1 Semantic Roles of Subjects in LV Clauses

There are 75 phonetically-realized subjects of LV clauses in the data, wherein 57% are theme, 35% are patient and only 8% are location. The semantic roles of subjects are definitely not a reliable criterion for the distinction between PV and LV (L. Huang 1993, 1995a, b, inter alia). The determinants are associated with modality, tense, aspect, animacy, and even lexical idiosyncrasy (S. Huang 2005). Similarly, as shown in Figure
5-2, theme and patient are major semantic roles carried by the subject in CV clauses, and beneficiary, the most characteristic semantic role of the subject in CV clauses, is in the minority.\textsuperscript{71}

In terms of phonetically-realized nominal expressions, among the whole 829 finite clauses, there are 131 clauses containing GTOPs (15.8%), 93 clauses containing GFOCs (11.2%), 78 clauses containing nominative clitics (9.4%), 193 clauses containing obliques (23.3%), 183 clauses containing objects (22.1%), and 363 clauses containing subjects (43.8%). These are illustrated in Table 5-2. The above 6 nominal expressions have 64 kinds of possible combinations ($\sum_{i=0}^{6} \binom{6}{i} = 2^6 = 64$). Only 30 of them are attested in the data, which are the 30 bars shown in Figure 5-3. All the descriptions in this section only represent the raw data. Null arguments will be added later in the analysis of inferential statistics since Squliq Atayal is a pro-drop language.

<table>
<thead>
<tr>
<th>Table 5-2 Manifestation of Syntactic Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Percent</td>
</tr>
</tbody>
</table>

\textsuperscript{71} S. Huang (2005) suggested that some cases of CV prefix $s$- should be analyzed as a causative marker. We will leave out this possibility in the present thesis since this issue may be associated with lexical idiosyncrasy or with diachronic lexico-semantic change.
5.2.2 Research Design

The present statistical analysis of topicality aims to discover what information each syntactic role conveys and the relationship between topicality, syntactic roles and voice alternation in Squliq Atayal. Although topicality is an abstract discourse property, there is a consensus in discourse analysis and functional syntax that it can be captured by the statistical measurements proposed in Givón’s series of works (Cf. 1983a, b, 1994, 1995, among others). As described in Chapter 2, one of them is RD and the other is TP. RD measures the extent of givenness—that is, how old the information is—by the distance between the previous occurrence of a nominal and its current occurrence in a clause. The shorter RD is, the more given information this nominal carries. TP measures the extent of aboutness—that is, how significant the information is—by the time of repetition of a nominal within the next 10 clauses following its present occurrence. The more TP is, the more important the information this nominal carries. The measure of RD and TP is “finite clause” because finite clauses code propositions, which combine

---

72 The abbreviations used in this figure are as the followings: V = verb; T = GTOP; P = predicative NPs; N = nominative clitics; G = oblique; O = object; S = subject.
concepts (i.e. words) into complete information.

Givón’s framework has three weaknesses. First, the measure “finite clause” is treated as interval. Interval data are more useful in statistics, but as Field (2009) reminds, “To say that data are interval, we must be certain that equal intervals on the scale represent equal differences in the property being measured (p.9).” It is rather debatable whether the amount of information carried by one finite clause equals that carried by another. In terms of the amount of syllables or words, most of linguists agree that finite clause is not a variable of equal interval: there are few finite clauses having the same amount of syllables or words in naturalistic data. Second, Givón’s serial works only look at the numbers of average and percentage. Average is easily affected by deviant values (outliers) and percentage cannot tell the extent of the difference between groups (Hatch & Lazaraton 1991, Larson-Hall 2010). However, the existence of outliers is highly possible since every nominal expression within the last finite clause in a text has no repetition at all, the value zero caused by which may make a great impact especially on the statistical results generated from averages. Therefore, the better way to avoid the first two problems is to adopt the chi-square test, which measures frequency and is not affected by outliers, as S. Huang and Tanangkingsing (2011) did. The chi-square test is inferential statistics, able to show whether the difference is statistically significant or not. Last, Givón’s framework only takes the intralinguistic context into account, ignoring the extralinguistic context. In linguistic fieldwork, all collected naturalistic texts are more or less related to the extralinguistic context since the informant is just in front of the researcher and whatever he or she says is definitely associated with the present context. Therefore, this extralinguistic factor is taken into consideration in the process of coding data.

Coding in the present analysis concerns syntactic roles and two topicality-related measurements, RD and TP. Only five syntactic roles are studied: GTOP, GFOC, subject, object, and oblique. Nominal expressions with these syntactic roles are coded with the
numbers from 0 to 4 respectively. RD is measured by the amount of finite clause between a nominal expression and its previous occurrence in discourse. The definition of previous occurrence contains all kinds of nominal expressions sharing the same referent, including null argument, pronouns, full NPs, and so forth. If there is a pair of identical nominal expressions but their referents are different, they will be viewed as not having any coherent relationship unless these two expressions refer to concept instead of entities. The maximum range is 10 finite clauses, starting from the one where the current nominal expression exists. The content of sub-context, like dialogues in a story, will be counted in the RD range. If there is no previous occurrence but the current nominal expression belongs to the superset-subset relationship with any nominal expression in the previous discourse, or the current nominal expression refers to some entity in the extralinguistic context, these nominal expressions will be regarded as inferable information, given a special number 11. Based on the value of each nominal expression, they are divided into groups: new (0), short (1-2), and long (3-10 & 11). With regard to TP, it is measured by the amount of repetition, to wit, how many times the referent recurs within the next 10 clauses following its present occurrence. The definition of repeated occurrence is just like that of previous occurrence of RD, and so is the maximum range. Again, if there is no repetition in the following discourse, but there is any nominal expression with the superset-subset relationship with the current nominal expression, the TP value is the number 11. Again, there are three groups: no (0), few (1-2, 11), and more (3-10). The unit used in the above two measurements is finite clause. Because tense/aspect marking is fused with voice markers in Atayal, the difference between finite clauses and non-finite clauses in this language consists in whether they are free to be affixed with different voice markers.

The independent variables are syntactic roles (GTOP, GFOC, subject, object, and oblique) and voices (AV vs. NAV). The dependent variables are RD and TP. All of the

73 This categorization is slightly modified from S. Huang (2002).
chi-square tests have a basic null hypothesis that independent variables make no difference on the frequency distribution of dependent variables. By rejecting this null hypothesis with 95% confidence interval, which means that every 100 identical experiment will lead to 95 times of the same result, it is claimed that there is a statistically significant difference in the frequency distribution of dependent variables, which may results from the effect of independent variables.

5.2.3 RD: Results

Syntactic roles are the independent variables in this chi-square test. There are five categories: GTOP, GFOC, subject, object, and oblique. Their RD is the dependent variable, which is divided into three categories: short (1-2 finite clauses), long (3-10 finite clauses or inferable information), and new (no previous occurrence). This ordering represents three levels of givenness: more given information, less given information and new information. The chi-square test will prove whether their frequency distributions have any significant difference. The result is shown in Table 5-3 below:
Table 5-3 *Referential Distance*

<table>
<thead>
<tr>
<th>Continuity</th>
<th>short</th>
<th>long</th>
<th>new</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>21</td>
<td>30</td>
<td>27</td>
<td>78</td>
</tr>
<tr>
<td>(%)</td>
<td>26.9</td>
<td>38.5</td>
<td>34.6</td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>(%)</td>
<td>13.0</td>
<td>52.2</td>
<td>34.8</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>249</td>
<td>148</td>
<td>44</td>
<td>441</td>
</tr>
<tr>
<td>(%)</td>
<td>56.5</td>
<td>33.6</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td>35</td>
<td>76</td>
<td>80</td>
<td>191</td>
</tr>
<tr>
<td>(%)</td>
<td>18.3</td>
<td>39.8</td>
<td>41.9</td>
<td></td>
</tr>
<tr>
<td>Oblique</td>
<td>121</td>
<td>97</td>
<td>20</td>
<td>238</td>
</tr>
<tr>
<td>(%)</td>
<td>50.8</td>
<td>40.8</td>
<td>8.4</td>
<td></td>
</tr>
</tbody>
</table>

($\chi^2=162.258$, df = 8, $p < .00$)

The $p$-value at the bottom of this table is below .00, much lower than .05. The number .05 is an important indicator in humanities and social sciences, which means that the probability of the confirmation of the null hypothesis is 5%. In other words, the rejection of the null hypothesis is up to 95%. Recall that the null hypothesis always suggests that independent variables have no effect on dependent variables. Thus, if the confidence to reject this null hypothesis is more than 95%, it is highly possible in statistics that the independent variables do affect dependent variables. Turn back to the above table, the $p$-value of which indicates that different syntactic roles (independent variable) do influence the distribution of RD values (dependent variable).

The values marked with boldface and green color in Table 5-3 are higher than our expectation—to wit, the values manifesting the influence of syntactic roles. In addition, the highest percentage within each category is marked with red color. The above table
shows several tendencies. First, GTOPs tend to bear less given information and even new information and so do GFOCs. The only distinction between them is the percentage of more given information, which is up to 26.9% of the former while only 13% of the latter. This dovetails the discovery in Rau (2000) that preverbal elements (so-called preverbal subjects in her paper) are topic discontinuity. The indistinguishable givenness of GTOPs and GFOCs may be partly due to the fact that GTOPs consist of not only argument topics but non-argument topics like adverbial topics as well and partly due to that even argument topics are switch topics, the information of which is not necessarily intralinguistically or extralinguistically given (Cf. Chapter 3). Second, subjects tend to convey more given information. Third, the information of objects tends to be new, not occurring in previous discourse, while that of obliques tends to be more given.

If the data are split into AV and NAV, the results are in Table 5-4. The \( p \)-values from AV clauses, NAV clauses and the whole data are all under .05, which indicates that there are statistically significant differences on the frequency distribution. Again, the numbers marked with boldface and green color are values higher than expectation, and the highest percentage within each category is marked with red color. In addition, those framed numbers are where the differences between Table 5-3 and Table 5-4 lie. The frequency distribution of AV clauses is the same as that of the whole. That is to say, preverbal elements, including GTOPs and GFOCs, and objects convey new or less given information while subjects more given information. As for NAV clauses, the information carried by preverbal elements (i.e. GTOPs and GFOCs) are new and less given information, indistinguishable from AV clauses. Obliques are also like what is shown in the whole: they convey more given information. What is intriguing is: although subjects are prone to bear given information, like what is shown in the whole, the new information carried by them is higher than expectation within all the categories.

\footnote{In order to make the chi-square test applicable, the objects of ditransitive NAV clauses are removed from this table.}
Put differently, subjects of NAV clauses are characterized by combining two extremes of topicality.

Table 5-4 *Voice* *Referential Distance*

<table>
<thead>
<tr>
<th>Continuity</th>
<th>short</th>
<th>long</th>
<th>new</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic/Focus</td>
<td>11</td>
<td>18</td>
<td>17</td>
<td>46</td>
</tr>
<tr>
<td>(%)</td>
<td>23.9</td>
<td>39.1</td>
<td>37.0</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>125</td>
<td>57</td>
<td>10</td>
<td>192</td>
</tr>
<tr>
<td>(%)</td>
<td>65.1</td>
<td>29.7</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td>31</td>
<td>71</td>
<td>77</td>
<td>179</td>
</tr>
<tr>
<td>(%)</td>
<td>17.3</td>
<td>39.7</td>
<td>43.0</td>
<td></td>
</tr>
<tr>
<td><strong>NAV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic/Focus</td>
<td>14</td>
<td>24</td>
<td>18</td>
<td>56</td>
</tr>
<tr>
<td>(%)</td>
<td>25.0</td>
<td>42.9</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>124</td>
<td>91</td>
<td>34</td>
<td>249</td>
</tr>
<tr>
<td>(%)</td>
<td>49.8</td>
<td>36.5</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>Oblique</td>
<td>119</td>
<td>97</td>
<td>19</td>
<td>235</td>
</tr>
<tr>
<td>(%)</td>
<td>50.6</td>
<td>41.3</td>
<td>8.1</td>
<td></td>
</tr>
</tbody>
</table>

(AV: $\chi^2 = 115.154$, df = 4, p < .00; NAV: $\chi^2 = 27.611$, df = 4, p < .00; Total: $\chi^2 = 69.488$, df = 4, p < .00)

The idealization of information manifestation concerning givenness is presented below (New = new information; Given\(+\) = less given information; Given\(++\) = more given information):
What is interesting is that AV and NAV have a great difference with regard to the arrangement of postverbal information: AV clauses start from new/less given information to more given one whereas NAV clauses from more given information to new/more given information.

5.2.4 TP: Results

This section turns to the statistical study on TP, studying how syntactic roles (the independent variable) affect TP (the dependent variable). Like what is done in the analysis of RD, syntactic roles are divided into five categories: GTOP, GFOC, subject, object, and oblique. TP is divided into three categories: more (3-10 repetitions), few (1-2 repetitions or inferable repetition), and no repetition. The more repetitions a nominal expression has in subsequent finite clauses, the more important it is to the whole discourse. Based on the chi-square test, syntactic roles do influence the frequency distribution of TP values, as presented in Table 5-5.
Table 5-5 *Topic Persistence*

<table>
<thead>
<tr>
<th>Persistence</th>
<th>more</th>
<th>few</th>
<th>no</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>22</td>
<td>39</td>
<td>17</td>
<td>78</td>
</tr>
<tr>
<td>(%)</td>
<td>28.2</td>
<td>50.0</td>
<td>21.8</td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>9</td>
<td>12</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>(%)</td>
<td>39.1</td>
<td>52.2</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>161</td>
<td>144</td>
<td>136</td>
<td>441</td>
</tr>
<tr>
<td>(%)</td>
<td>36.5</td>
<td>32.7</td>
<td>30.8</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td>36</td>
<td>48</td>
<td>107</td>
<td>191</td>
</tr>
<tr>
<td>(%)</td>
<td>18.8</td>
<td>25.1</td>
<td>56.0</td>
<td></td>
</tr>
<tr>
<td>Oblique</td>
<td>108</td>
<td>73</td>
<td>57</td>
<td>238</td>
</tr>
<tr>
<td>(%)</td>
<td>45.4</td>
<td>30.7</td>
<td>23.9</td>
<td></td>
</tr>
</tbody>
</table>

($\chi^2 = 81.199$, df = 8, p < .00)

The $p$-value of the above table is below .05, indicating that the result is statistically significant. Those values higher than our expectation are marked with boldface and green color, and the highest percentage within each category is marked with red color.

GTOPs tend to be repeated for one or two times. In other words, GTOPs have a scope over one or two finite clauses most of the time.\(^{75}\) This is not unexpected since a sentence holding to the discourse topic may have a different topic of its own (sentence topic). GTOPs are merely syntactically-marked switch topics, the function of which is to attract attention to a nominal expression that is not currently focused on. GFOCs have a great resemblance to GTOPs in that the majority of them have small scope over one or

\(^{75}\) The scope of GTOPs is smaller than that of subjects in Squiq Atayal. This is different from Japanese: the scope of GTOPs is bigger than that of subjects, instead (Hind 1983). One of the possible reasons is that GTOPs in Squiq Atayal include those non-argument topics like adverbial topics, left-dislocated topics, etc., but GTOPs in Japanese don’t. These non-argument topics do not last long in the discourse. Specifically speaking, the difference between these two languages may be due to the fact that GTOPs in them have different discourse functions.
two finite clauses. With regard to subjects and obliques, they both have higher degree of importance in discourse, usually with a wide scope over more than two finite clauses; reversely, objects do not have high degree of discourse importance, having no repetition in the sequent clauses most of the time.

Again, the data are divided into AV and NAV clauses. It is tested whether syntactic roles affect TP in these two groups. The results are shown in Table 5-6. The p-values of AV clauses, NAV clauses and the whole are all below .05, which indicates the statistical significance of this test. Values higher than expectation are marked with boldface and green color, and those values different from what we see in the analysis of the whole data are framed in this table. The highest percentage within each category is marked with red color.

Table 5-6 Voice*Topic Persistence

<table>
<thead>
<tr>
<th>Persistence</th>
<th>more</th>
<th>few</th>
<th>no</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic/Focus</td>
<td>17</td>
<td>18</td>
<td>11</td>
<td>46</td>
</tr>
<tr>
<td>(%)</td>
<td>37.0</td>
<td>39.1</td>
<td>23.9</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>90</td>
<td>61</td>
<td>41</td>
<td>192</td>
</tr>
<tr>
<td>(%)</td>
<td>46.9</td>
<td>31.8</td>
<td>21.4</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td>31</td>
<td>44</td>
<td>104</td>
<td>179</td>
</tr>
<tr>
<td>(%)</td>
<td>17.3</td>
<td>24.6</td>
<td>58.1</td>
<td></td>
</tr>
<tr>
<td>NAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic/Focus</td>
<td>14</td>
<td>34</td>
<td>8</td>
<td>56</td>
</tr>
<tr>
<td>(%)</td>
<td>25.0</td>
<td>60.7</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>71</td>
<td>93</td>
<td>95</td>
<td>249</td>
</tr>
<tr>
<td>(%)</td>
<td>28.5</td>
<td>33.3</td>
<td>38.2</td>
<td></td>
</tr>
<tr>
<td>Oblique</td>
<td>108</td>
<td>71</td>
<td>56</td>
<td>235</td>
</tr>
<tr>
<td>(%)</td>
<td>46.0</td>
<td>30.2</td>
<td>23.8</td>
<td></td>
</tr>
</tbody>
</table>
(AV: $\chi^2 = 63.892$, df = 4, $p < .00$; NAV: $\chi^2 = 37.984$, df = 4, $p < .00$; Total: $\chi^2 = 25.712$, df = 4, $p < .00$)

Compared with the general pattern in Table 5-5, the above table shows that preverbal elements in AV transitive clauses, that is, GTOPs and GFOCs, tend to stay longer in the discourse while it is not the case in NAV clauses. This difference may be due to the subject-only constraint on topicalization (Cf. Chapter 3): GTOPs/GFOCs in AV clauses are prone to be agent while those in NAV clauses are prone to be non-agent. Agent is always under the spotlight of discourse line. Subjects and objects in AV clauses are in the complementary distribution in terms of information status: the former are very important in discourse while the latter are unimportant in discourse. So are subjects and obliques in NAV clauses, but the pattern is reverse: subjects are unimportant in the discourse while obliques are very important. In terms of information status, subjects in AV clauses and obliques in NAV clauses look alike, and so do objects in AV clauses and subjects in NAV clauses. This phenomenon coincides with S. Huang & Tanangkingsing (2011). If we take a closer look at the values of subjects in NAV clauses, however, we will discover that they are dispersed wider than those of objects in AV clauses.

The results are demonstrated in the following idealized representations ($\text{About}_{(++)} =$ more TP; $\text{About}_{(+)} =$ few TP; $\text{About}_{(-)} =$ no TP):

<table>
<thead>
<tr>
<th>(107)</th>
<th>Topic</th>
<th>Focus</th>
<th>$V_{(AV)}$</th>
<th>Object</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
<td>$\downarrow$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
</tr>
<tr>
<td>$\text{About}_{(++)}$</td>
<td>$\text{About}_{(++)}$</td>
<td>$\text{About}_{(-)}$</td>
<td>$\text{About}_{(++)}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The above representations are abstract but helpful to capture some general pictures of the manifestation of informational aboutness/importance in Squiliq Atayal. Whether in AV or NAV clauses, preverbal elements, GTOPs and GFOCs, have discourse importance to some extent. What is more intriguing is the reverse arrangement between AV and NAV clauses. AV clauses match the typological universal that subjects tend to have more repetitions in subsequent clauses than objects, which is discovered in Amharic (Gasser 1983), English (Brown 1983, Wright & Givón 1987), Javanese (Subiyanto 1998), Korean (Kwak 1994), and so forth. NAV clauses however do not behave like passive clauses at all in that the oblique actors still play an important role in the discourse. According to Thompson (1997), the distinction between core arguments and obliques lies upon their givenness (i.e. RD), identifiability and tracking (i.e. TP): core arguments tend to be given, identifiable and important whereas obliques new, unidentifiable and unimportant.

These idealized representations are not without exception. It is not an absolute requirement that objects of AV transitive clauses should be unimportant, without being repeated in the subsequent clauses. The following example is one of the exceptions:

(109) ungat qu’ squiliq nyux m-lahang ngasal qani laga.
    Neg Nom person Aux.Prog AV-take.care.of house this Par

‘There is no one who is taking care of this house.’ <N>

76 Givón’s (1979a) study on English showed that “the function of passive sentences in language is to code sentences in the context in which the non-agent is more topical. This automatically means that the agent is less topical in a passive sentence and the fact that it gets removed ([demoted]) from the subject slot—the one which usually coincides with the topic—is an obvious means of achieving this end (p.57).”
The above sentences are adjacent in the same text. In (109), the main clause is composed of the negative predicate *ungat* and the subject *squliq* 'person', the latter of which is further modified by a relative clause. The object of the relative clause *ngasal qani* ‘this house’ is a definite noun phrase conveying given information, and most important of all, it is repeated as a GTOP in the following discourse in (110). It is obvious that givenness (RD) and importance (TP) are the only two determinants of voice alternation in Squliq Atayal. First of all, we have to keep in mind that importance, the assumed concept behind TP, is merely the speaker’s prediction of the subsequent discourse in the current clause where the nominal expression exists. Whether the prediction is born out and manifested by the TP value in the final discourse product is not guaranteed. The fact that the sentence in (110) topicalizes *ngasal qani* ‘this house’ indicates this noun phrase is foregrounded from the background. Put differently, the direction of the subsequent discourse line also influences the TP values. In addition, in (109), the event manifested by the auxiliary *nyux* is progressive, and the main verb *mlahang* ‘take care of’ is prefixed with *m*- , the AV marker, which also indicates strong degree of agentivity or volitionality of the subject. To be more specific, AV clauses excel in expressing dynamic actions, so in order to satisfy this event semantics, it does not matter whether the object is important or not in the following discourse.

In the same vein, obliques in NAV clauses may not be as topical as the idealized representations show. The exceptional example is illustrated below:

(111) knbaq=nya’ ga, ini’ ltx.
acquaintance=3SGen Top Neg AV.bark
‘(As for) its acquaintances, (it) does not bark.’ <N>
The sentences in (111) and (112) are adjacent in one narrative about the informant’s dog at home. Because the discourse topic is very explicit, none of the grammatical subjects in these two examples, which definitely refer to the dog, are phonetically-realized. Take a close look at the first clause of (112), the subordinate clause, and it shows that the main verb s’angun ‘shout down’ is suffixed with the PV marker -un, which indicates that the null subject (i.e. the dog) is the patient of this event. The null agent of this event however is not explicit at all. It is not mentioned in the preceding sentences and not repeated in the subsequent sentences. It cannot be clearly inferred through extralinguistic context, either. This NAV clause functions exactly like passive clauses in English.77

5.2.5 Summary

The tendencies of information manifestation in Squiql Atayal can be summarized below:

**AV transitive clauses:**
- GTOPs and GFOCs tend to convey new or less given information and they are more important in the subsequent discourse;
- Subjects tend to convey more given information and they are more important in the subsequent discourse;
- Objects tend to convey new or less given information and they are unimportant in the subsequent discourse.

**NAV clauses:**
- GTOPs and GFOCs tend to convey new or less given information and they are less important in the subsequent discourse than those of AV transitive

---

77 Givón (1979a) proposed three categories of agentless passives in English: one in which agent is recoverable from following discourse, another in which agent is not recoverable in the direct discourse context but via a more general body of background knowledge, and the other in which the recoverability of agent is associated with lexical meaning. Example (159) belongs to the second category.
clauses;
- Subjects convey diverse types of information, in particular more given and new information, and they are slightly inclined to be unimportant in the subsequent discourse;
- Obliques tend to convey more given information and they are more important in the subsequent discourse.

These tendencies can be demonstrated in the following idealized representations:

<table>
<thead>
<tr>
<th>(113)</th>
<th>New/Given(+), New/Given(+)</th>
<th>New/Given(+), Given(++)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New/Given(+), New/Given(+)</td>
<td>New/Given(+), Given(++)</td>
</tr>
<tr>
<td></td>
<td>Topic</td>
<td>Focus</td>
</tr>
<tr>
<td></td>
<td>⇕</td>
<td>⇕</td>
</tr>
<tr>
<td></td>
<td>V(AV)</td>
<td>Oblique</td>
</tr>
<tr>
<td></td>
<td>⇕</td>
<td>⇕</td>
</tr>
<tr>
<td></td>
<td>About(+/++)</td>
<td>About(+/++)</td>
</tr>
<tr>
<td></td>
<td>About(+/++)</td>
<td>About(+/++)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(114)</th>
<th>New/Given(+), New/Given(+)</th>
<th>New/Given(+), New/Given(++)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New/Given(+), New/Given(+)</td>
<td>New/Given(+), New/Given(++)</td>
</tr>
<tr>
<td></td>
<td>Topic</td>
<td>Focus</td>
</tr>
<tr>
<td></td>
<td>⇕</td>
<td>⇕</td>
</tr>
<tr>
<td></td>
<td>V(NAV)</td>
<td>Oblique</td>
</tr>
<tr>
<td></td>
<td>⇕</td>
<td>⇕</td>
</tr>
<tr>
<td></td>
<td>About(+)</td>
<td>About(+)</td>
</tr>
<tr>
<td></td>
<td>About(+)</td>
<td>About(+)</td>
</tr>
</tbody>
</table>

Linguists must be careful in making use of these tendencies since there are always exceptions despite their low proportion.

Prior to the next section, it has to be noted that the statistical study of this section has its limitation. Quantitative studies are restricted by their sampling. Chapter 2 has pointed out that these data were collected from limited native speakers. The data of an ideal linguistic corpus must be sampled from different genres, themes, registers and people of different ages, genders, socio-economic backgrounds (Hunston 2002, McEnery & Wilson 2001, Meyer 2002, Yang 2002). All of the data must be balanced in
that all variables have the same amount of tokens (i.e. texts of the same length). Thus, corpus linguists have confidence to proclaim that their data are representative enough of the whole language. In reality, this kind of sampling is far beyond the reach of the fieldwork on minority languages. What field linguists can do is adopt the assumption that all native speakers of one language share the same grammatical knowledge and tentatively ignore the sociolinguistic variables behind.

5.3 Information Chaining

This section returns to the analysis of information structure by means of qualitative method by investigating how the information of two adjacent sentences is connected. It will demonstrate that information chaining in Squiql Atayal is correlated with topicality, syntactic roles and voice alternation.

5.3.1 Thematic Progression

With regard to how information is passed down from one sentence to another, Daneš (1974) proposed “thematic progression”, which is “the choice and ordering of utterance themes, their mutual concatenation and hierarchy, as well as their relationship to the hyperthemes of the superior text units (such as the paragraph, chapter, …), to the whole text, and to the situation (p.114).” Here (utterance) theme is neither given information at the sentence level (i.e. sentence/clause topics) nor the first element of a clause, a term used in the Prague School or SFG. Instead, it is the information expressed by nominal, which can be either given or new. Thematic progression is close to what Chafe (1976, 1979) called information flow. There are three types of thematic

---

78 Even by adopting such a strict sampling method, as Atkins & Rundell (2008) points out, corpora still have biases in word frequency. It is not unexpected because corpus linguists cannot control the use of vocabulary and syntactic structures in any text. Sounds, words and structures are not equally distributed over all texts. In order to compensate this insufficiency, it is encouraged to establish one corpus as large as possible.
progression: simple linear, constant theme, and derived theme. They are illustrated respectively in (115), (116), and (117) from Daneš (1974:118-119):

(115) Simple Linear

\[
\begin{align*}
T_1 &\rightarrow R_2 \\
T_2 &= R_3 \\
T_3 &= R_4
\end{align*}
\]

(116) Constant Theme

\[
\begin{align*}
T_1 &\rightarrow R_2 \\
T_1 &= R_3
\end{align*}
\]

(117) Derived Theme

\[
\begin{align*}
T_1 &\rightarrow R_1 \\
T_2 &\rightarrow R_2 \\
T_3 &\rightarrow R_3
\end{align*}
\]

In Daneš (1974), theme (T) is given information (i.e. sentence/clause topic) while rheme (R) is new information. The above three types of thematic progression are actually three ways of connecting information across the sentence boundary. The first one in (115), simple linear, indicates that the second sentence takes the new information of the preceding sentence as its own starting point (theme/sentence topic). Put differently, the rheme of the first sentence changes the direction of information flow by introducing a new piece of information. The second type in (116), constant theme, is the information connection made by connecting several sentences with coreferential themes. The direction of information flow is not changed at all in this case. Every theme of the following sentence is a continuing topic. As for the last type in (117), derived theme, it represents the indirect connection between the information of two sentences.

A similar idea was given different names in different works, such as “topic chain” in Dixon (1972) and Tsao (1979, 1990), which focuses on how a topic connects a series
of sentences. Erteschik-Shir & Lappin (1979) and Erteschik-Shir (1988) termed the first type as “dominance chaining” and the second type as “topic chaining”, and Erteschik-Shir (1997) termed these three types as “focus chaining”, “topic chaining”, and “restrictive focus chaining” respectively. Fries (1993) called them “linear thematic progression”, “Theme iteration”, and “a progression with derived Themes”, respectively. In Smith (2003), these three types are termed as “focus-topic chaining”, “topic chaining”, and “unchained”. Linguistic terms like theme, rheme, topic and focus have many definitions, which already led to much unnecessary confusion. In order to avoid confusion and to emphasize the information connection between sentences, thematic progression is called “information chaining” in the present thesis. The three types of chaining are labeled as “new-given chaining”, “given-given chaining”, and “inferable chaining”. Only the first two types are the focus of this section since inferable chaining is associated with non-linguistic world knowledge and reasoning, the mechanism of which is beyond the scope of the present thesis.

Nevertheless, it is not enough only to take information status into account because syntactic roles wield a considerable influence. As described in Smith (2003), “[t]he topic phrases…are all surface structure subjects…This is the position that recurs strikingly across the text passages examined for this study (p.246).” Accordingly, in what follows, the discussion on information chaining will be based on four combinations of voices: AV-AV adjacency, AV-NAV adjacency, NAV-AV adjacency, and NAV-NAV adjacency. Under each adjacency, it will be shown how voice alternation (AV and NAV) and syntactic roles (subject and non-subject) affect information chaining. There are 16 hypothetical patterns (4 Adjacency Types*2 Arguments*2 Chaining Types), but not all of them are attested by the data of the present thesis.

79 The application of thematic progression has been extended far beyond the scope of syntactic studies into discourse/text analysis (Fries 1983, Scinto 1983, inter alia). This application is sidestepped in the present thesis.
5.3.2 Adjacencies (I)

5.3.2.1 AV-AV Adjacency

When two AV sentences are adjacent in discourse, this adjacency is termed as AV-AV adjacency. In AV sentences, the most important arguments are subjects and objects. The following discussion will look at subjects first, to investigate whether the topicality of subjects, new or given, has any effect on information chaining, and then to examine objects of different topicality as well. All of the patterns presented from this section are all illustrated with figures, wherein new information is marked with red color, given information blue color, and the information chaining is visualized with an arrow.

In AV-AV adjacency, a piece of new information introduced by the subject of an AV sentence may be passed down to the subject of the next AV sentence, as illustrated in Figure 5-4 and instantiated in (118) and (119).

Figure 5-4 New-Given Chaining of AV-AV Subject
(118) raral ga, m-aki’ beh yatux, kahul yatux qu’ past Top AV-live near high.mountain AV.originate high.mountain Nom mama’ qasa.
uncle that
‘A long time ago, that uncle lived in high mountains, and came from high mountains.’ <N>

(119) a, kani ru’ m-aras hango Ø ra.
Par AV.walk and AV-take container Ø Par
‘(He) walked and took a cylindrical container.’ <N>

These two sentences are extracted from the beginning of a text. The first sentence introduces the character of the story through its subject, mama’ qasa ‘that uncle’. The second sentence continues the story line by placing this piece of information in the subject position as well. This type of information chaining is common when an actor is introduced into the discourse. However, there is no case in the data that the lately-introduced information in the first sentence is repeated as the object of the second AV sentence.

In contrast, given information can be easily passed down from the subject of an AV sentence to either the subject or the object of another AV sentence. The cases of the former are much more than those of the latter. This pattern is illustrated in Figure 5-5 and exemplified in (120) and (121).

Figure 5-5 Given-Given Chaining of AV-AV Subject
(120) nanu’ yasa m-wah Ø, ru’ m-wah shzi’ m-aki’ slaq qani Ø.
Par thus AV-come Ø and AV-come AV-move AV-live wet.field this Ø
‘Well, thus, (they) came, and (they) immigrated to live in this wet field.’ <N>

(121) cyux kya Ø ga, ata k-m-yap Ø.
exist there Ø Top once AV-catch Ø
‘While (the frog) is being there, once (I) tried to catch (it)…’ <N>

The example in (120) demonstrates that a piece of given information is shared by the subjects of two adjacent AV sentences. The given information is manifested as null subjects marked by Ø in this example. In (121), the subject of the first sentence, marked by the first Ø, refers to frogs, which is a piece of given information already introduced in previous sentences, and the second sentence is connected with the first sentence by repeating this piece of information in the object position, marked by the second Ø.

If a new-given chaining starts with the object of an AV sentence, it only can end at the object of the second AV sentences in AV-AV adjacency like Figure 5-6.

Figure 5-6 New-Given Chaining of AV-AV Object (I)

(122) ngilis ga m-ita yaqih, m-ita yaqih hi laga.
AV-howl Top AV-see devil AV-see devil one Par
‘When (the dog) howls, (it) sees a devil, sees a devil.’ <N>
All attested cases of this pattern, however, involve iterations of the same event, as (122). The object *yaqih* ‘devil’ is introduced into the discourse for the first time, and it is repeated again with the verb *mita* ‘see’ in the second sentence. Strictly speaking, it is not the objects that connect the two adjacent AV sentences. There is one marginal case in (123), which can be viewed either as a coordinated predicate or as sequent sentences. If this example is analyzed as two independent AV sentences, of course, the object *utux* ‘ghost’ connects their information flow.

In addition, there is no case of the new-given chaining between the object of an AV sentence and the subject of another AV sentence. Squliq Atayal is greatly different from English in that in the former the object conveying new information cannot chain AV-AV adjacency while in the latter it can chain two adjacent active sentences. In other words, the development of a text by rheme in Baker (1992) does not work in Squliq Atayal. It will be discussed in Chapter 6 that they do not correspond to presentational focus in i-structure.

Given information, even conveyed by the object of an AV sentence, is less restricted in information chaining. Put differently, the given-given chaining with the object of an AV sentence as the head can connect two AV sentences no matter whether its tail is the subject or the object of the second AV sentence, as illustrated in Figure 5-7.
(124) cyugan ga, m-ita 'byan=nha', m-aki' balay 'byan=nha' uzi.
third one Top AV-see bed=3PGen AV-exist really bed=3PGen too
'(As for) the third one, (I) see their bed. Their bed really exists as well.' <N>

(125) nanak qutux yumin sugi <n>aki' qani.
only one Yumin Sugi <Perf> live this
‘Only Yumin Sugi’s family lived in this place.’ <N>

(126) lha' uzi ga iyat m<n>aki' qani.
3PNeu also Top Neg AV<Perf> live this
‘(Speaking of) them, (they) had not lived in this place either.’ <N>

The example in (124) is a given-given chaining between the object of an AV sentence and the subject of another AV sentence. The noun phrase ‘byan-nha’ ‘their bed’ occurs twice: one functions as the object and the other functions as the subject. As for (125) and (126), they are connected by the object qani ‘this’ as the object of the first AV sentence and qani ‘this’ as the subject of the second one.

5.3.2.2 AV-NAV Adjacency

AV-NAV adjacency means that there are two adjacent sentences in discourse wherein the first one is an AV sentence and the second is a NAV sentence. Although subjects are more topical than objects in AV sentences, it is the opposite in NAV
sentences: obliques are more topical than subjects. This tendency has been presented in the first half of this chapter, and undoubtedly it affects information chaining.

When the subject of an AV sentence conveys new information, this information can easily connect the next NAV sentence no matter whether it ends as a subject or an oblique. This is illustrated in Figure 5-8.

Figure 5-8 New-Given Chaining of AV-NAV Subject

(127) ungat qu’ bingxiang, s-on=nha’ bingxiang, theluw. Neg Nom fridge call-PV=3PObl fridge mainland ‘There are no fridges. They call fridges, the mainlanders.’ <N>

(128) si’ kta-y la-ga, s-m-wal qu’ k’aki’=maku’ la. if look-PV Par-Top AV-agree Nom deceased.grandmother=1SGen Par ‘Before long, my deceased grandmother agreed.’ <N>

(129) s-biq=nya’, ake, k’aba’=maku’ qu’ k’aya’=maku’ la. CV-espouse=3SObl Par deceased.father=1SGen Nom deceased.mother=1SGen Par ‘She espoused my deceased mother to my deceased father.’ <N>

The example in (127) is a case of AV-NAV adjacency. The subject bingxiang ‘fridge’ is a piece of new information introduced into discourse via the subject of the AV sentence, and this information is repeated in the subject position of the second sentence. Another example is in (128) and (129), where a new character k’aki’-maku’ ‘my deceased
grandmother’ as the subject of the AV sentence recurs as a pronominal oblique ‘nya’ ‘she’.

In the same vein, the given information conveyed by the subject of an AV sentence can be passed to the subject or the oblique of a NAV sentence without any difficulty, as shown in Figure 5-9 and the examples from (130) to (133).

Figure 5-9 Given-Given Chaining of AV-NAV Subject

(130) kt-an l-ga qyanux Ø lozi.  
see-LV Par-Top AV.live Ø again  
‘After being seen, (the ghost) came back to life.’ <N>

(131) ay, ini’ balay baq-un k-m-ut qu’ utux.  
Excl Neg really can-PV AV-kill Nom ghost  
‘Alas, the ghost cannot really be killed.’ <N>

(132) “utux qlw-an” m-ha qu’ yaki’ sa.  
ghost float-LV AV-say Nom aunt that  
‘That aunt thought, “a floating ghost.”’ <N>

(133) hwag-an=nya’ kya lozi.  
swear-LV=3SObl then again  
‘Then she swore at (it) again.’ <N>

The subject in (130) is an invisible argument referring to the ghost mentioned in the
previous discourse, and *utux* ‘ghost’ occurs again as the subject of a NAV sentence in (131). The subject in (132) *yaki*’ *sa* ‘that ghost’ conveys given information likewise. This information is repeated as the oblique of a NAV sentence in (133).

As for the object of an AV sentence, if it conveys new information, this information cannot connect adjacent sentences. There is no attested example in the data. In contrast, if it conveys given information, this information is able to connect the two sentences of the AV-NAV adjacency by ending as either the subject or the oblique of a NAV sentence, as shown in Figure 5-10. This is instantiated in examples from (134) to (137).

**Figure 5-10 Given-Given Chaining of AV-NAV Object**

(134) *nanu’a, sami’ nanak kbalay baliq qasa uzi.*
Par Par 1PNeu only AV.make iron that too
‘Well, we alone made the bullet.’ <N>

(135) *nanu’a, bhy-an=myan qu’ baliq qa rru.*
Par Par hit-LV=1PObl Nom iron that Par
‘Well, we hit that bullet.’
‘Well, that bullet was hit by us.’ (lit.) <N>
(136) hiya’ nanak qu’ nyux s-m-ru’ balay la, sami’ ake, tayal 3SNeu only Nom Aux.Prog AV-support really Par 1PNeu Par Atayal qani ga.
  this Par
  ‘The one who is really supporting us, these Atayal people, is only her.’ <N>

(137) y-an ni’ sami’ qu’ bulak=myan qani.
  resemble-LV Obl 1PNeu Nom tribe=1PGen this
  ‘We are like our tribe.’ <N>

The first two sentences are connected through the given information conveyed by baliq qa(sa) ‘that bullet’, which functions as the object in (134) and as the subject in (135). The last two sentences exemplify the given-given chaining ending as the oblique. The free pronoun sami’ ‘we/us’ in (137) is an oblique because it is preceded by the oblique case marker ni’.80

5.3.3 Adjacencies (II)

5.3.3.1 NAV-AV Adjacency

The NAV-AV adjacency is a pair of two sentences, the first of which is a NAV sentence while the second of which is an AV sentence. If a new-given chaining starts from the subject of a NAV sentence, it always ends at the object of an AV sentence, not at its subject, as illustrated below:

---

80 Why does this sentence not use the oblique clitic -myan? Long pronouns always have an emphatic meaning. As the English translation of (137) shows, “us” is emphasized. Using the oblique clitic does not bring out this discourse-pragmatic effect.
Figure 5-11 New-Given Chaining of NAV-AV Subject (I)

(138) kt-an=nha’, son mha, **yaqi utux** akuma.
see-LV=3PObl so-called evil ghost demon
‘They saw so-called evil ghosts, demons.’
‘So-called evil ghosts, demons, were seen by them.’ (lit.) <N>

(139) nanu’a, kun uzi ga, m-ita Ø uzi.
Par Par 1SNeu too Top AV-see Ø too
‘Well, (as for) me too, (I) saw (ghosts) as well.’ <N>

The above two sentences are adjacent in a text. The subject **yaqi utux** ‘evil ghost’ in the
NAV sentence (138) is a piece of new information, which connects the next sentence
through the object in the AV sentence (139). The new information conveyed by the
subject of a NAV sentence can be passed to the subject of an AV sentence through
topicalization, as shown in Figure 5-12.
Figure 5-12 New-Given Chaining of NAV-AV Subject (II)

(140) m-wah ga, ras-un=nha’ qu’ qutux niqan.
AV-come Top take-PV=3PObl Nom one hunting.group
‘When (they) came, they brought a hunting group.’
‘When (they) came, one hunting group was brought by them.’ (lit.) <N>

(141) qutux niqan qasa ga, cingay balay.
one hunting.group that Top AV many really
‘(Speaking of) that hunting group, (they) were really many.’ <N>

These two sentences are adjacent in one text. The subject *qutux niqan* ‘one hunting group’ in (140) is a piece of new information. It is repeated as the GTOP in the next sentence.

In contrast, when the subject of a NAV sentence conveys given information, this information can easily connect adjacent sentences. It can be realized as the subject or the object of an AV sentence. The examples in (142) and (143) exemplify the given-given chaining between the subject of a NAV sentence and the subject of an AV sentence; the given information carried by *baliq qa* ‘that bullet’ recurs as the null subject in (143). The example in (144) instantiates the given-given chaining between the subject of a NAV sentence and the object of an AV sentence.
(142) nanu’ a, bhy-an=myan qu’ baliq qa rru.
Par Par hit-LV=1PObl Nom iron that Par
‘Well, we hit that bullet.’
‘Well, that bullet was hit by us.’ (lit.) <N>

(143) tapaq balay Ø l-ga, …
fat very Ø Par-Top
‘When (it) is very fat…’ <N>

(144) nanu’ a, s-’agal baliq. nanu’ a, sami’ nanak kbalay baliq qasa uzi.
Par Par CV-take iron Par Par 1PNeu only AV.make iron that too
‘Well, (we) took iron. Well, we alone made the bullet.’ <N>

The obliques conveying new information are in the decided minority, and none of them connect NAV-AV adjacency in the attested examples. In other words, the new-given chaining initiated by obliques is unattested in the data. Conversely, the obliques conveying given information are able to connect two adjacent sentences easily, as shown in Figure 5-14.
Given-Given Chaining of NAV-AV Oblique

(145) soliq qasa ga, baq-un=**myan** kbalay kwara.’
woodgun that Top know-PV=1PObl AV.make all
‘Speaking of woodguns, we all know how to make it.’ <N>

(146) nanu’ a, m-usa i, s-m-’ung squ’ qhoniq Ø.
Par Part AV-go Par AV-chop Loc tree Ø
‘Well, (we) go to chop in the forest.’ <N>

(147) gukoh, s-on=**myan** na’ **neiwan.** tehuk=ta’ **gukoh** laga…
Gukoh call-PV=1PObl Obl Neiwan AV.arrive=1PNom Gukoh Par
‘(Speaking of) Gukoh, (it) is what we call Neiwan. After we arrive at Gukoh…’<N>

The first two sentences exemplify the given-given chaining between the oblique of a NAV sentence and the subject of an AV sentence. The oblique -**myan** ‘we/us’ in (145) is undoubtedly a piece of given information, and then this information is further passed down to the null subject Ø in (146). The sentence in (147) is an instance of the information chaining between the oblique of a NAV sentence and the object of an AV sentence because the given information conveyed by **neiwan** and **gukoh** refer to the same town.
5.3.3.2 NAV-NAV Adjacency

The last type of adjacency is a combination of two adjacent NAV sentences. The new-given chaining starting from the subject of a NAV sentence can only end at the subject of another NAV sentence, as shown in the following figure:

Figure 5-15 New-Given Chaining of NAV-NAV Subject

(148) kt-an=nya’ qu’ yaya’=maku’ ro.
   see-LV=3SObl Nom mother=1SGen Par
   ‘He saw my mother.’
   ‘My mother was seen by him.’ (lit.) <N>

(149) ki’an=nya’ szy-on balay qu’ yaya’=maku’.
   perhaps=3SObl love-PV really Nom mother=1SGen
   ‘Perhaps he really loved my mother.’ <N>

The above sentences are adjacent in a text. The first one introduces the new information yaya’-maku ’ ‘my mother’ into the discourse, and then this information is repeated as the subject of another NAV sentence. This pattern is common, but there is no case of the new-given chaining between the subject of a NAV sentence and the oblique of another NAV sentence.

Nevertheless, when the information conveyed by the subjects in NAV clauses is
given, the restriction mentioned above disappears. The given-given chaining can exist between the subject of a NAV sentence and either the subject or the oblique of another NAV sentence, like Figure 5-16.

Figure 5-16 Given-Given Chaining of NAV-NAV Subject

(150) nanu’a, kt-an zipun Ø l-ga, wal i kut-an tayal nanak Ø. Par Par see-LV Japanese Ø Par-Top Aux.Pt Par kill-LV Atayal only Ø
‘Well, after (he) was discovered by the Japanese, (he) was killed by the Atayal people themselves.’ <N>

The example in (150) is extracted from a story about an Atayal murderer. This murderer has been introduced into the discourse before this sentence, so it is not phonetically-realized in these two sentences. The murderer is the subjects of these two NAV sentences. The given-given chaining between the subject of a NAV sentence and the oblique of another NAV sentence is instantiated in (151) and (152). The speaker herself has been mentioned twice: the nominative clitic -ku’ in (151) and the oblique clitic -maku’ in (152).

(151) …s-kaki’=ku’=nha’ ngasal.
CV-exist=1SNom=3PObl house
‘…They gave birth to me in the house.’ <N>
Although there is no attested new-given information chaining starting from the oblique in the NAV-AV adjacency in Section 5.3.3, it does exist in the NAV-NAV adjacency. Since the oblique of a NAV sentence tends to convey given information, there is only one case of new-given information chaining, as shown in Figure 5-17 and exemplified in (153) and (154).

Figure 5-17 New-Given Chaining of NAV-NAV Oblique

(153) hango hiya’ ga, giw-an cbilan.
   container Emp Top resemble-LV lunch.box
   ‘(Speaking of) cylindrical containers, (they) are like lunch boxes.’ <N>

(154) cbilan qasa ga, sy-an mami’ Ø la.
   lunch.box that Top load-LV rice Ø Par
   ‘(Speaking of) those lunch boxes, (they) are loaded with rice.’ <N>

The first sentence introduces the entity cbilan ‘lunch box’, serving as an oblique, into the discourse for the first time. In the second sentence, the entity cbilan ‘lunch box’ is repeated as the GTOP, then further serving as the grammatical subject. The existence of this case shows that the new information conveyed by obliques is not important to the
information flow so topicalization is a necessary syntactic strategy to foreground it.

In contrast, as the majority, obliques with given information can connect two adjacent NAV sentences, like Figure 5-18. The tail of this given-given chaining can be either the subject or the oblique of a NAV sentence.

Figure 5-18 Given-Given Chaining of NAV-NAV Oblique

(155) ssy-an=\textit{myan} squ’ baliq qa.
\hspace{1cm}load-LV=1PObl Loc iron that
‘We load (it) with that bullet.’ <N>

(156) wah-an=\textit{myan} i, wah-an pqelun son mha.\textsuperscript{81}
\hspace{1cm}go-LV=1PObl Par go-LV trigger so-called
‘We pull, well, pull the so-called trigger.’
‘The so-called trigger is pulled, well, pulled by us.’ (lit.) <N>

(157) nanu’ s-ngugu’ knan Ø? hwag-an=\textit{myan} Ø.
\hspace{1cm}what CV-scare 1SNeu Ø swear-LV=1SObl Ø
‘For what did (he) scare me? I swore at (him).’ <N>

The examples in (155) and (156) exemplify the case of information chaining between the obliques of two adjacent NAV sentences. The oblique -\textit{myan} ‘we/us’ has two

\textsuperscript{81} The noun \textit{pqelun} ‘trigger’ is derived from the LV form of the verb \textit{hoqil} ‘die’ with the causative prefix \textit{p}-. The word in this story refers to a component of the Atayal’s traditional woodgun so the morphological information is omitted from the gloss.
occurrences. This pattern is frequent in the data. The example in (157) instantiates the information chaining between the oblique of a NAV sentence and the subject of another NAV sentence, which are both null arguments marked by $\emptyset$.

### 5.3.4 Summary

Information chaining in Squilq Atayal is a complex discourse phenomenon correlated with several factors: topicality (given vs. new), syntactic roles (subject vs. non-subject) and voice alternation (AV vs. NAV). Given information is unrestricted to chain two adjacent sentences no matter what voices these sentences are marked with and what syntactic roles the head and the tail of the chaining play. In contrast, the syntactic role of a piece of new information determines whether it has the ability to connect the next sentence and in what way the two adjacent sentences can be chained. Briefly speaking, it is more restricted by syntactic factors at the sentence level. The patterns of new-given chaining in Squilq Atayal are summarized below:

Table 5-7 New-Given Chaining in Squilq Atayal

<table>
<thead>
<tr>
<th>Head \ (New)</th>
<th>Tail \ (Given)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV subject</td>
<td>AV subject</td>
</tr>
<tr>
<td>NAV subject</td>
<td>NAV subject</td>
</tr>
<tr>
<td>NAV oblique</td>
<td>NAV oblique</td>
</tr>
<tr>
<td>AV object</td>
<td>AV object</td>
</tr>
<tr>
<td>NAV subject</td>
<td>AV object</td>
</tr>
<tr>
<td>NAV subject</td>
<td>NAV subject</td>
</tr>
<tr>
<td>NAV oblique</td>
<td>unattested</td>
</tr>
</tbody>
</table>

Not all new information is able to initiate information chaining even though new
The new information conveyed by subject is almost unrestricted while the new information conveyed by object and oblique is highly restricted and even unattested. In order to pass the new information conveyed by them to another sentence, it is imperative to employ topicalization or repetition. Information chaining is definitely not syntax-free. It needs the collaboration between different components of a language.

5.4 Conclusion

This chapter presents two pieces of discourse evidence: one concerns the quantitative analysis of topicality and the other concerns the qualitative analysis of information chaining. The former analysis shows that although there is no distinction between AV and NAV clauses in terms of the topicality of GTOPs and GFOCs, this distinction does exist for the topicality of postverbal arguments: subjects are more topical than objects in AV clauses, while subjects are less topical than obliques in NAV clauses. The information conveyed by subjects in NAV clauses is rather diverse, with both tokens of given information and those of new information. The latter analysis of information chaining shows that it is easier for given information to initiate information chaining than new information. The information chaining initiated by new information is restricted by syntactic roles and voice alternation. These two pieces of discourse evidence pin down the core of the present thesis again that information structure and other syntactic components cooperate with each other in Squilq Atayal. Their cooperation is manifested in the mapping between discourse functions in i-structure and grammatical functions in f-structure, the theoretical issue discussed in the following three chapters.

82 The importance of new information has been researched under the studies on rheme in SFG research (Halliday 1967a, 1985). As Baker (1992) states, “[t]he rheme is what the speaker says about the theme. It is the goal of discourse. As such, it is the most important element in the structure of the clause as a message because it represents the very information that the speaker wants to convey to the hearer. It is the rheme that fulfills the communicative purpose of the utterance (p.122).”
6.

Correspondences (I):
Unmarked Clauses

Neither can embellishments of language be found without arrangement and expression of thoughts, nor can thoughts be made to shine without the light of language.

--Cicero

6.1 Overview

In addition to the discourse evidence in Chapter 5, the present thesis provides more evidence from two syntax-oriented tests in the beginning of this chapter: question-answer tests and thetic judgments. The question-answer tests show that the placement of new information in the question part affects discourse-pragmatic appropriateness of a grammatical sentence; this piece of evidence is in conflict with the quantitative analysis of topicality especially in terms of subject. The discussion on thetic judgments manifests that this tests should not be treated as a clear-cut criterion for differentiating subject-prominence languages and topic-prominence languages.

Both of the syntax-oriented tests and the discourse evidence, no matter whether they are qualitative or quantitative evidence, facilitate the identification of what discourse functions in i-structure each of the core arguments, subject, object and oblique, correspond to. Based on the discourse functions identified from core arguments, we can observe the general mapping between i-structure, f-structure and c-structure, which further benefits explaining empirical puzzles about the typological classification
between f-structural languages and i-structural languages, the essence of multiple voices, and the formation of the unmarked word order VOS.

6.2 Syntax-Oriented Tests

In this section, two syntax-oriented tests, question-answer tests and thetic judgments, are employed to examine the mapping between f-/c-structure (i.e. syntax) and i-structure (i.e. discourse-pragmatics).

6.2.1 Question-Answer Tests

The distinction between given and new information is the core of research on information structure since most of well-known languages are sensitive to this distinction. The mini discourse context of question-answer pairs helps narrow down the seemingly messy discourse-related syntactic phenomena into rather simple tidy patterns because the answer part consists of two parts, one part conveying given information identical with the information told in the question part while the other part conveying new information (also called semantic focus in Jackendoff (1972)) filling the communicative gap between interlocutors. Whether this bipartition affects syntax stands out in the answer part. In English, it has no bearing on syntax, with only prosody affected. Squiliq Atayal however is more sensitive to this bipartition.

The interrogative sentences in Squiliq Atayal usually employ pseudo-cleft constructions with fronted interrogative words as the nominal predicate of matrix clauses in the sentence-initial position; most important of all, if fronted interrogative words are arguments, they must be the grammatical subjects of the embedded nominal clauses, which trigger the subject-verb agreement on thematic roles in this language. If

---

83 Question-answer tests are the groundwork for the Question Under Discussion models of information structure, in which either explicitly-asked or implicitly-understood questions are the basic building blocks of information structure (Velleman & Beaver 2016). For instance, the discourse-tree in Büring (2003) has discourse move as a node, the relationship between two discourse moves as a branch, and answering move as a terminal node.
interrogative arguments are not grammatical subjects, they must stay in-situ, remaining at the normal position of its grammatical functions. In other words, in-situ interrogative arguments are not pseudo-clefted. Details of the formation of interrogative sentences will be discussed in Chapter 8. In question-answer tests, interrogative words in the question part are placed in three different positions, serving as three different grammatical functions: pseudo-clefted (preposed) subject, (in-situ) object and (in-situ) oblique actor.

The following sentences from (158) to (161) constitute the first test of the question-answer pair, in which the question part contains a pseudo-clefted (fronted) interrogative word:

(158) Q: **ima’** (qu’) **wal** m-ihiy tali’?
who Nom Aux.Pt AV-hit Tali
‘Who hit Tali?’ <E>

(159) A: **watan** (qu’) **wal** m-ihiy tali’.
Watan Nom Aux.Pt AV-hit Tali
‘The one who hit Tali was Watan.’ <E>

(160) A:# **wal** m-ihiy tali’ (qu’) **watan**.
Aux.Pt AV-hit Tali Nom Watan
‘Watan hit Tali.’ <E>

(161) A: **wal** bhy-an watan **qu’** tali’.
Aux.Pt hit-LV Watan Nom Tali
‘Tali was hit by Watan.’ <E>

If the answer part is made with a pseudo-cleft construction as in (159), it will be discourse-pragmatically perfect. This answer consists of two parts: one is the pseudo-clefted nominal **watan** (marked in green), which conveys the information unknown to the addressee, and the other is the embedded clause (qu’) **wal mihiy tali’**
‘the one who hit Tali’ (marked in red), which offers the information shared with the question part and with the addressee. If the answer part is not a pseudo-cleft construction, like the sentence in (160), wherein the given information becomes the predicate while the new information becomes the grammatical subject, this answer part will not be discourse-pragmatically appropriate at all even though it is syntactically grammatical while isolated in grammaticality judgments. If the voice marking of the answer part is altered to be LV, as presented in (161), wherein the new information becomes an oblique actor while the rest of the sentence still conveys given information, the answer part will still manifest discourse-pragmatic appropriateness.

The second test of question-answer pair is composed of the sentences from (162) to (166), as shown below:

(162) Q: cyux k-m-ut **nanu’** qu’ tali’?  
      Aux.Prog AV-kill what Nom Tali  
      ‘What is Tali killing?’ <E>

(163) A: cyux k-m-ut **ngta’** qu’ tali’.  
      Aux.Prog AV-kill chicken Nom Tali  
      ‘Tali is killing a chicken.’ <E>

(164) A: * **ngta’** cyux k-m-ut qu’ tali’.  
      chicken Aux.Prog AV-kill Nom Tali <E>

(165) A: **ngta’** (qu’) cyux kut-an (na’) tali’.  
      chicken Nom Aux.Prog kill-LV Obl Tali  
      ‘What is being killed by Tali is a chicken.’ <E>

(166) A: * cyux kut-an (na’) tali’ (qu’) **ngta’**.  
      Aux.Prog kill-LV Obl Tali Nom chicken <E>

The question part in (162) has the in-situ interrogative word **nanu’** ‘what’, serving as the
object of the AV verb *kmut* ‘kill’. This interrogative sentence does not involve pseudo-clefting. If the answer part replaces the interrogative word with new information in the object position, like *ngta’* ‘chicken’ in (163), it is discourse-pragmatically appropriate without a doubt. The second answer in (164) is ungrammatical because as will be shown in Chapter 8, grammatical objects cannot be pseudo-clefted. If the voice marking of the answer part is altered to be LV and the new information is pseudo-clefted, as shown in (165), the answer will be discourse-pragmatically appropriate. By contrast, as shown in (166), if the new information is not pseudo-clefted, remaining in situ as the subject, the answer part will be discourse-pragmatically inappropriate even though it is syntactically grammatical.

The last test of question-answer pairs starts with a question sentence containing an interrogative word as the oblique actor, like *ima’* ‘who’ in (167).

(167) Q: wal kut-an **ima’** qu’ ngta’ qu ngani?
   Aux.Pt kill-LV who Nom chicken this
   ‘By whom was this chicken killed?’ <E>

(168) A: wal kut-an **watan** qu’ ngta’ qu ngani.
   Aux.Pt kill-LV Watan Nom chicken this
   ‘This chicken was killed by Watan.’ <E>

(169) A:* watan wal kut-an qu’ ngta’ qu ngani.
   Watan Aux.Pt kill-LV Nom chicken this <E>

(170) A:# wal k-m-ut ngta’ qu’ watan.
   Aux.Pt AV-kill chicken this Nom Watan
   ‘Watan killed this chicken.’ <E>
New information, as *watan* (marked in green) in (168), can substitute the oblique interrogative word of the question part, forming a discourse-pragmatically appropriate answer part. The second answer sentence in (169) is ungrammatical because obliques cannot be pseudo-clefted. If the answer part of the same proposition is expressed with an AV sentence, like (170), it will result in discourse-pragmatic inappropriateness since the new information *watan* is located in the subject position. However, there will be no problem if the new information is pseudo-clefted, removed from the subject position, as illustrated in (171).

The above three tests of question-answer pairs show that whether an answer sentence is discourse-pragmatically appropriate depends upon the position of new information. To be more specific, the new information of the answer part cannot be placed in the (sentence-final) subject position; that is, it cannot be a subject. As long as this restriction is violated, even grammatically correct sentences cannot be the answer part.

### 6.2.2 Thetic Judgments

By the means of the Brentano-Marty theory in logic, Kuroda (1972) makes a distinction between two kinds of propositions, one containing a notional subject and a notional predicate (categorical judgments), while the other containing only a notional predicate (thetic judgments). To be more specific, the formation of categorical judgments is composed of two acts: one is the recognition of the referent which is made the notional subject, and the other is the affirmation/denial what the predicate expresses about the notional subject. In contrast, thetic judgments have only the act of affirming/denying the relation expressed by the notional predicate. These are
Sentence (172) contains the proposition with both a notional subject and a notional predicate; therefore, the formation of its proposition is to identify the notional subject from the previous discourse first, as in the a-part of (173), and then to identify the notional predicate, as in the b-part of (173). However, the formation of the proposition of sentence (174) is only concerned with the identification of the notional predicate running (i.e. the b-part of (175)) since the grammatical subject a dog cannot be identified from the previous discourse as presented in the a-part of (175).

Kuroda’s (1972) distinction is purely based on semantics, but it gives an implication that a language which structurally realizes thetic judgments exactly like categorical judgments is a subject-prominent language because this language favors grammatical subject so much that it is structurally realized even though the notional subject does not exist (É. Kiss 1995). For instance, English is a typical example of subject-prominent language:

(176) (a) **It** is raining. (Impersonal) (É. Kiss 1995:7)
(b) **There** is a dog in the room. (Existential)
(c) **A dog** came into the room. (Presentative)
(d) **All dogs** like bones. (Universal)
Impersonal, existential, presentative and universal sentences are structural substantiation of thetic judgments, which are supposed to have no notional subject. However, they do have grammatical subjects in English. As for those languages in which the structural substantiation of thetic judgments is different from that of categorical judgments, they are topic-prominent languages. É Kiss (1995) stated that “in topic-prominent languages the (primary) syntactic predication structure is always directly mapped on a notional predication relation, whereas in subject-prominent languages this is not the case (p.13).”

Based on Kuroda’s (1972) dichotomy, it is intriguing to know whether our target language is a topic-prominent language or a subject-prominent language. The answer comes from the following thetic judgments in Squilq Atayal:

177) mosa qwalax Ø. (Egerod 1980:540)
   Aux.Fut AV.rain Ø
   ‘It will rain.’

178) nyux m-hlaqiy Ø. (Egerod 1999:72)
   Aux.Prog AV-snow Ø
   ‘It is snowing.’

179) m-aki’ qutux kinholan ratal mga…
   AV-exist one elder.person before Par
   ‘There was an old woman in the past…’ <N>

180) nanu’a, ratal ga, m-aki’ qu’hato.
   Par Par before Top AV-exist Nom wild.pigeon
   ‘Well, once upon a time, there was a wild pigeon.’ <N>
The examples in (177) and (178) are impersonal sentences with weather verbs, in which there is no subject following the predicate *mosa’ qwalax* ‘will rain’ or *nyux mhlauy* ‘be snowing’ respectively. This is different from the subject-prominent language English, which must have a dummy subject *it* in impersonal sentences. Impersonal sentences manifest the topic-prominence of Squiliq Atayal. Nevertheless, there are grammatical subjects in existential, presentative and universal sentences, like *qutux kinholan* ‘one elder person’ (179), *hato* ‘wild pigeon’ in (180), *zipun* ‘Japanese’ in (181) and *kwara’ utux* ‘all gods’ in (182). The last three kinds of sentences are like canonical sentences with sentence-final subject. These are typical characteristics of subject-prominent languages. Among the above four tests of thetic judgments, one shows that Squiliq Atayal is a topic-prominent language whereas the other three show that Squiliq Atayal is a subject-prominent language. Why does this discrepancy occur?

One possible explanation is that there is an invisible dummy subject in impersonal sentences of Squiliq Atayal, if one insists that the thetic-judgment tests make a clear demarcation of subject-prominent languages from topic-prominent languages, as claimed by É Kiss (1995). Although this explanation maintains the consistency and

---

84 The intervention of the nominative case *qu* ‘between *qutux* ‘one’ and *zipun* ‘Japanese’ can be analyzed in two ways. The first possible analysis is that this example is a case of quantifier-floating, a common phenomenon in languages. The second one is that *qutux* ‘one’ is the predicate of *zipun* ‘Japanese’ and *qutux qu* ‘zipun’ is an internally-headed relative clause, the existence of which is attested in (Liu 2004a, b, 2005).

85 The gloss of *utux* in Rau (1992) is ‘god’, not ‘ghost’. The original translation of this sentence is ‘In the past all the ghosts were angry.’ The Atayal people believe that after death, a person’s ghost will continue to protect his/her descendants, like a god. The translation in Rau (1992) is not wrong, but to be consistent, the word *utux* is glossed as ‘ghost’ in the present thesis.
validity of thetic judgments, it is an ad hoc stipulation since this invisible dummy subject exists nowhere except in impersonal sentences. What is more, the inconsistency of thetic judgments still appears in topic-prominent languages, as illustrated in the following Mandarin Chinese sentences:

(183) Ø zhengzai xia yu.
    Ø Prog fall rain
    ‘It is raining.’

(184) Ø you yi-zhi gou zai na-ge fangjian.
    Ø have one-Cl dog in that-Cl room
    ‘There is a dog in that room.’

(185) Ø you yi-zhi gou zoujin na-ge fangjian.
    Ø have one-Cl dog come in that-Cl room
    ‘A dog came into that room.’

(186) gou dou xihuan gutou.
    dog all like bone
    ‘All dogs like bones.’

It is a well-known fact, since Li and Thompson (1976), that Mandarin Chinese is a typical topic-prominent language. The sentences from (183) to (185) show that there is no grammatical subject in impersonal, existential and presentative sentences. However, in universal sentences like (186), there is a grammatical subject gou ‘dog’. These sentences prove that thetic judgments do not work well even on typical topic-prominent languages. The discrepancy has been pointed out by Xu (2005[2002]).

In order to account for this, it is possible to propose another language-particular

---

86 More details on these three sentence types can be referred to Chao (1968), Chu (2010), Li & Thompson (1981), Liu et al. (1996[1983]), and Tang (1979[1977]), inter alia. Unlike traditional analyses, Tsai (2003) analyzes you “have” as an unselective binder, not a modal/verb, in existential and presentative sentences. If his research is on the right track, it is highly possible to analyze yi-zhi gou “one dog” as a grammatical subject. Then, thetic-judgment tests will look more useless in determining whether a language is topic-prominent or not.
stipulation, just like what is proposed for Squliq Atayal in the beginning of this paragraph. Honestly speaking, this kind of language-particular stipulation merely wipes the question under the carpet rather than solves it. So does the complete abandonment of thetic judgments, as Xu and Liu (2007) did because of their adherence to the typological classification proposed by Li and Thompson (1976), since most of subject-prominent European languages are diagnosed as topic prominent languages under thetic judgments.

Alternatively, contrary to É Kiss’s (1995) idea that subject-prominent languages and topic-prominent languages are typologically exclusive, the present thesis agrees to the idea that there is a continuum between radical subject-prominent languages and radical topic-prominent languages, which results from the interactions between i-structure and other syntactic modules. These thetic-judgment tests should not be treated as homogeneous criteria. The inconsistent outcome of thetic-judgement tests is due to recessive factors from the respective mini discourse contexts of these four sentences, which should be delved into one by one rather than lumped together. In other words, only by recognizing their internal divergences can linguists really tell what the syntactic interactions with information structure are in each language and make more subtle typological comparisons. Being eager to adopt or to reject É Kiss’s classification will deprive us of the opportunity to observe how linguistic systems operate internally.

Of these four thetic-judgment sentences, impersonal sentences do not indicate any referent but only an event especially in languages where there are a group of weather verbs, like Squliq Atayal and English. The former adopts different strategy from the latter. In Squliq Atayal, non-existent information is not syntactically substantiated. This strategy complies with both the iconicity principle that morphosyntactic structures have to be as close to the conceptual structures as possible and the economy principle that morphosyntactic structures should make the least effort to fulfill the most
communicative functions (Haiman 1983, 1985). Under the LFG framework, the c-structure of this kind of languages is more sensitive to i-structure than f-structure, the details of which will be discussed in the next section. Thus, as illustrated in (177) and (178), there is no argument in an impersonal sentence except the weather verb itself. The other strategy is adopted in English and many other European languages. Owing to syntactic requirements, the subject position has to be filled with a dummy subject. Although this violates iconicity and economy, it achieves the regularity of morphosyntactic realization—that is, every sentence in this kind of languages always has a subject-predicate configuration. Put it in LFG terminology, c-structure in languages of this kind is more sensitive to f-structure than i-structure. As for languages like Mandarin Chinese, in which there are no weather verbs, impersonal sentences do have referents, i.e. the entities (like snow and rain) occurring in weather phenomena. Their positions follow the language-particular constraints on the interactions between syntax and information structure.

Existential and presentative sentences are closely related in terms of their mini discourse contexts. These two sentences serve to introduce a new participant into the information flow (Dryer 2007, Quirk et al. 1972, 1985, Ward, Birner & Huddleston 2002). Erteschik-Shir (2007) claims that they have implicit scene-setting topics, which are exactly the tempo-spatial scene where the event exists. As shown in (179), (180) and (181), the new participant is placed in the subject position in both existential and presentative sentences of Squliq Atayal, a fact running afoul of what we found in the tests of question-answer pairs. In addition, the new participant introduced through these two sentence types must be important in the global discourse context or the local

87 Most of the time, the iconicity principle competes with the economy principle since complex concepts need complex syntactic structures to express while simple structures economize the effort to complete communication (Haiman 1983, 1985). The competition between them is manifested in the ranking competition between faithfulness constraints and markedness constraints in the OT (Cf. Kager 1999, Prince & Smolensky 1993). The harmony between iconicity and economy is not impossible, as shown in coordination reduction (Haiman 1983).
discourse context where it exists (Givón 1993). Otherwise, addressers do not bother to introduce a new participant. This also proves that the subject position in this language is closely related with important information, i.e. the information with [+ABT]. The same mapping between grammatical subject and important information no matter that information is given or new is also reported in other predicate-initial languages, such as Ojibwa (an American language) in Tomlin and Rhodes (1992). Remind that the mini discourse context of existential and presentative sentences is different from that of question-answer pairs. On the one hand, it is reasonable to state that these two sentence types have their own mapping pattern between syntax and information structure. For instance, they both require grammatical subjects conveying new information, counter to the general requirement of given information. On the other hand, these two sentence types display the commonality of grammatical subject in this language. That is, grammatical subjects in Squliq Atayal convey important information, which has been pinpointed in Chapter 4. Details and LFG mechanism will be discussed in the next section.

As for universal sentences, their communicative function varies with discourse contexts. Although É. Kiss (1995) claims there is no act of identifying referents in universal sentences, the universally quantified subject may serve either given or new information, depending upon what context the whole sentence occurs. This is illustrated below:

(187)  Q: What did you learn from today’s lesson?  
       A: All dogs like bones.

(188)  Q: What do you know about dogs?  
       A: All dogs like bones.

Of the above two question-answer pairs, the first one has all dogs as a part of the new
information, which is the whole sentence itself, while the second one has *all dogs* as given information, which merely repeats what the first speaker says, and the rest of it conveys new information. Put it simply, universal sentences are not relevant to any specific mapping between syntax and information structure. This sentence pattern is not helpful to elucidate the interactions between them.

In conclusion, this section demonstrates that the application of the thetic-judgment tests to differentiate subject-prominent languages and topic-prominent languages always brings about discrepancies. The inconsistency of their outcome manifests that there is a continuum between radical topic-prominent languages and radical subject-prominent languages and that the clear-cut typological dichotomy between these two types, as proposed in É. Kiss (1995), does not hold. These thetic sentences however provide linguists the opportunities to observe how syntax interacts with information structure since they represent different special mini discourse contexts. In addition, Squliq Atayal has characteristics of both subject-prominence and topic-prominence, seeming to stand between the two ends.

### 6.3 Arguments in Mapping

Based on both discourse evidence and syntax-oriented tests discussed earlier in the present thesis, this section will illustrate what discourse functions the three most important arguments in unmarked sentences, subject, object and oblique actor, convey respectively and how these correspondences are annotated to phrase structure rules.

#### 6.3.1 Subject

The question-answer tests in Section 6.2.1 demonstrate a rather neat fact that grammatical subject in Squliq Atayal obligatorily conveys given information, not permitting new information, no matter whether the answer sentence is an AV clause or a NAV one. In contrast, other arguments in unmarked clauses, that is, object and oblique
actor, are not restricted by whether or not the information is shared between the addresser and the addressee. In addition, the discourse evidence from information chaining in Chapter 5 shows that it is easier for the new information introduced by the subject of either an AV sentence or a NAV sentence to connect discourse than for the new information introduced by other arguments. It is manifest that information conveyed by grammatical subjects is important to subsequent discourse.

Despite the neatness of qualitative evidence, the discourse reality demonstrated in the statistical studies of Chapter 5 is complicated. In terms of RD, the measurement of givenness, 95% of the subjects in AV clauses and 86% of the subjects in NAV clauses convey given information, with short or long RD. In terms of TP, the measurement of aboutness, 79% of the subjects in AV clauses and around 62% of the subjects in NAV clauses convey important information, with more or few TP. The statistical outcome is not as neat as that of the question-answer tests. If the two essential discourse features at i-structure, givenness and aboutness, are taken into consideration simultaneously, the following distributional patterns will come out:

Table 6-1 Discourse Features of Vt Subjects

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>AV(%)</th>
<th>NAV(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>137 (76.5)</td>
<td>134 (53.8)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>33 (18.4)</td>
<td>81 (32.5)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>5 (2.8)</td>
<td>20 (8.0)</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>4 (2.2)</td>
<td>14 (5.6)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>179</td>
<td>249</td>
</tr>
</tbody>
</table>

As shown in the above table, it is not inviolable in the discourse reality that grammatical subjects have to convey CTTOP, with aberrant case up to 23.5% in AV clauses and 46.2% in NAV one, although the majority of grammatical subjects still meet this
requirement with the tokens of \([+\text{GIV}, +\text{ABT}]\) up to 76.5\% in AV clauses and 53.8\% in NAV ones.

The existence of the aberrant cases in quantitative evidence runs afoul of the question-answer tests.\(^8\) How can we give an account of the discrepancy between the qualitative evidence from the question-answer tests and the quantitative evidence from Givón’s method? Traditional syntacticians favor syntax-oriented tests, thinking of statistical aberrant cases negligible, whereas discourse analysts despise grammaticality judgments, emphasizing the importance of diversity. This bifurcation in methodology makes linguistics a blooming academic field lacking internal communication most of the time. Once again, as has been emphasized in the present thesis, some ideological and methodological bifurcations between linguistic schools are subject to compromise. The solution comes from whether we can really understand the nature of these two methods and then give them the best interpretations. The explanation of the present thesis is as follows: what the question-answer tests demonstrate is the default mapping between grammatical subject and CT TOP in Squliq Atayal while those seeming exceptions in quantitative studies prove the one-to-many mapping from subject to diverse discourse functions in discourse reality, which results from prototypes in cognition as well as cooperation in communication.

The above explanation can be supported through a closer look at the distribution of tokens in Table 6-1. It is intriguing that the distribution forms a gradience from the highest percentage of \([+\text{GIV}, +\text{ABT}]\) (76.5\% in AV clauses and 53.8\% in NAV clauses),

\(^8\) It had seldom been touched upon in the formalist approach that some clear-cut syntactic rules run afoul of linguistic reality—the absolute accuracy of one hundred percent seems to exist only in our abstract theories—until the emergence of OT. The OT approach solves this problem by proposing violable constraints and constraint ranking (Archangeli & Langendoen 1997, Kager 1999, Prince & Smolensky 1993, among others). The weakness of OT comes from the uneven distributional frequency and from the recessive discourse-pragmatic forces behind linguistic facts (Croft 2003). Take for example the default mapping of grammatical subjects in question. Since the majority of grammatical subjects convey given and important information, the constraints prohibiting subjects with new information or with unimportant information must be dominant, with great effect on the output patterns. As for the occurrence of those exceptional cases, it is attributed to the interference of more dominant constraints, whose effect can override that of the no-new/unimportant-information-for-subject constraint.
to the second high percentage of [+GIV, −ABT] (18.4% in AV clauses and 32.5% in NAV clauses), to the lower percentage of [−GIV, +ABT] (2.8% in AV clauses and 8% in NAV clauses), at last to the lowest percentage of [−GIV, −ABT] (2.2% in AV clauses and 5.6% in NAV clauses), instead of random or even distribution. In addition, not only transitive clauses but also intransitive clauses display similar distribution, as shown in Table 6-2: the highest percentage still goes to the group with [+GIV, +ABT] (60.3%) and the second highest one goes to the group with [+GIV, −ABT] (25.4%). The only difference is that the lowest percentage does not go to the group with [−GIV, −ABT] but the one with [−GIV, +ABT] though the percentage of both groups is below 10%. The majority of subjects in intransitive clauses still convey given information, up to 85.7%.

Table 6-2 Discourse Features of Vi Subjects

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>AV(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>190 (60.3)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>80 (25.4)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>19 (6.0)</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>26 (8.3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>315</td>
</tr>
</tbody>
</table>

According to Givón’s quantitative method, for either transitive clauses or intransitive ones, the information conveyed by grammatical subjects forms a hierarchy presented in (189). The information at the left-hand side of the hierarchy is prioritized as a grammatical subject in f-structure over those at the right-hand side. From the perspective of i-structure, the group with [+GIV, +ABT] at the leftmost hand has the most prototypical property of CTTOP, so it is the largest group among these four

---

89 Traugott & Trousdale (2010, 2013) distinguishes “gradience” from “gradualness”: the former refers to variation in synchronic grammar without generational discreteness while the latter refers to change over time (that is, in diachronic grammar) with generational discreteness.
categories. If there is interference from other factors, the other groups with non-prototypical properties on the right-hand side are backups. It is not difficult to realize why the group with \([+\text{GIV}, -\text{ABT}]\) is the largest among these backup candidates since it preserves the most essential characteristic of a topic, and since the importance of information is not obligatory in that a discourse usually has its last clause as the ending, in which no information can be viewed as important in the subsequent discourse.

\[(189) [+\text{GIV}, +\text{ABT}] > [+\text{GIV}, -\text{ABT}] > [-\text{GIV}, +\text{ABT}] > [-\text{GIV}, -\text{ABT}]\]

The concept of prototype in cognition lays the groundwork for the substitution of \(\text{CTTOP}\) with discourse functions containing similar discourse features. It is proposed that the real mechanism in i-structure is a conversion function called “koppa”, symbolized by \(Q\). It converts one discourse function into another one. The reasoning behind \(Q\) is simple. Native speakers are aware of and assume the existence of a particular correspondence between a discourse function in i-structure and one function in another representation. In the process of communication, the addresser makes use of this assumption between interlocutors to place another discourse function which does not belong to this correspondence into it, as long as the violation of the default mapping is not highlighted in morphosyntactic structures or in discourse, such as the mini discourse contexts of the question-answer pairs. The fundamental principle behind it is Grice’s (1975) “cooperative principle”. The idea originated from “presupposition” in Stalnaker (1973) and “pragmatic accommodation” in Lambrecht (1994). In the present case, Squiq Atayal has the conversion function in (190), in which the first part \((\alpha \in \text{DF} \land \tilde{C}_{i \to f}(\alpha) = \text{SUBJ})\) stipulates the condition that there is a discourse function \(\alpha\) in i-structure corresponding to subject (i.e. grammatical function \(\text{SUBJ}\) in f-structure) and the second part \((Q(\alpha) = \text{CTTOP})\) requires that \(\alpha\) become a continuing topic (i.e. discourse function \(\text{CTTOP}\) in i-structure) through the application of the conversion function \(Q\).
\( \alpha \in \text{DF} \land \check{c}_{i \rightarrow f}(\alpha) = \text{SUBJ} \Rightarrow Q(\alpha) = \text{CTTOP} \)

Put differently, when the information conveyed by a grammatical subject does not exactly fit into the prototypical properties of continuing topics, it is still regarded as a continuing topic via the communicative cooperation between interlocutors. The cognitive mechanism behind this is “spreading activation”, the (near-)simultaneous activation of closely related notes in particular usage events (Hudson 2010, Traugott & Trousdale 2013). Put plainly, when we are looking for the best fit in the process of communication, typically our choice has to be made in microseconds and we certainly have no time to consider all of possibilities. It always happens that we eventually pick out the second or the third best choice; that is, we simultaneously activate the nodes closely related to our target (the best choice) in the cognitive network.

Although the conversion function \( Q \) provides the flexibility of information insertion into i-structure from discourse-pragmatic contexts, it also erodes the default mapping between CT TOP in i-structure and \( \text{SUBJ} \) in f-structure. With the increase of tokens with non-prototypical properties, the original mapping type will lose its ground and concede to a new mapping over time, which leads to diachronic changes eventually. It is more possible for this erosion to happen to NAV clauses than AV clauses since actors (the subject of AV clauses), especially agents, are more inclined to be part of discourse topic in information flow while non-actors (the subject of NAV clauses) are not, which is manifested in the percentage of aberrant cases in Table 6-1. NAV clauses are more likely to apply the conversion function.

The case of the grammatical subjects conveying new and unimportant information, i.e. the aberrant cases, is exemplified in (191), which is excerpted from a text on traditional weapon.
(191) wah-an=myan i, wah-an pqlun son mha.
go-LV=1PObl Par go-LV trigger so-called

‘We pull, well, pull the so-called trigger.’

‘The so-called trigger is pulled, well, pulled by us.’ (lit.) <N>

It is a NAV sentence with *pqlun son mha* ‘so-called trigger’ as the subject. The component trigger is introduced into this text for the first time through this sentence, and it has no more mentions later. Although it conveys new and unimportant information in statistics, there is no denying that its information is closely related with the general discourse topic: a trigger is the most important component providing the physical force for arrows in a weapon. The close relationship between the information conveyed by this subject and the general discourse topic compensates for the non-prototypicality of being information conveyed by a subject, further facilitating the application of Q.

Based on both quantitative and qualitative evidence, the annotated phrase structure rules (192) and (193) are proposed for unmarked sentences in Squliq Atayal, the first one for lexical nominal while the second one for pronominal clitics. \(^90\) The annotations of these phrase structure rules in c-structure mark the correspondences between c-structure and f-structure, and c-structure and i-structure.

\(^90\) Although clitics and non-clitics have different phrase structure rules in c-structure, making a distinction between them does not change the fact that both clitics and non-clitics prototypically correspond to CTTOPs in i-structure, which is shown in Table (i). It is displayed that non-clitics in both AV clauses and NAV clauses still form a gradience of discourse features, a pattern exactly like what is presented while lumping clitics and non-clitics together. As for clitics, they even exhibit a stronger tendency toward CTTOPs in i-structure.

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>AV (%)</th>
<th>AV (%)</th>
<th>NAV (%)</th>
<th>NAV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Non-clitics</td>
<td>Clitics</td>
<td>Non-clitics</td>
<td>Clitics</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>94 (79.0)</td>
<td>43 (71.7)</td>
<td>117 (51.1)</td>
<td>17 (85.0)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>16 (13.4)</td>
<td>17 (28.3)</td>
<td>78 (34.1)</td>
<td>3 (15.0)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>5 (4.2)</td>
<td>0</td>
<td>20 (8.7)</td>
<td>0</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>4 (3.4)</td>
<td>0</td>
<td>14 (6.1)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>119</td>
<td>60</td>
<td>229</td>
<td>20</td>
</tr>
</tbody>
</table>
In (192), the sentence-final position at c-structure corresponds to both the grammatical subject in f-structure (indicated by \( \uparrow \text{SUBJ} = \downarrow \)) and a continuing topic in i-structure (indicated by \( \uparrow_i \text{CTTOP} = \downarrow \)) simultaneously. This topic must convey given and important information, which is stood for by the above annotations with fixed values \( \downarrow_i \text{GIV} = \text{c +} \) and \( \downarrow_i \text{ABT} = \text{c +} \). The subscript c marks the restriction of the values of discourse features givenness and aboutness. This restriction only restricts the general mapping between c-structure and i-structure, leaving the room of applying Q in i-structure itself. What the question-answer tests show is this default mapping. The phrase structure rule in (193) is basically like that in (192) except that it requires that the pronominal clitics should be nominative case (indicated by \( \downarrow \text{CASE} = \text{NOM} \)), the first or the second person (indicated by \( \downarrow \text{PERS} = \text{c 1\text{\textsuperscript{\textbf{v}} 2}} \)), and undoubtedly pronominal (indicated by \( \downarrow \text{PRED} = \text{‘PRO’} \)).

In addition, one of the most remarkable differences between the phrase structure rule in (192) and the one in (193) lies in \@DEFAULT (PROM, −) and \( \downarrow_i \text{PROM} = \text{c −} \). The symbol @ precedes the call of a template. An template in LFG is a label for a
functional description representing the generalization of a set of equations and constraints that describes linguistic structures (Asudeh 2012, Asudeh, Dalrymple & Toivonen 2008, Butt et al. 1999, Dalrymple, Kaplan & King 2004). The present thesis introduces a template for the default requirement of discourse features as defined below:

\[(194) \quad \text{@DEFAULT (attribute, value)} = \{(\uparrow_i \text{attribute}) = \text{value} \bigm| (\uparrow_i \text{attribute}) \neq \text{value}\}\]

The above template makes an overridable stipulation that an attribute in i-structure has a certain value unless there is a case of this attribute which has the opposite value. The attribute in the above template is a generalized name for various discourse features, such as PROM in the present case. Therefore, \text{@DEFAULT (PROM, −)} means that the discourse feature PROM has the negative value (−) by default unless there is an exceptional case. That is to say, the phrase structure rule (192) does not require the information must be nonprominent while the phrase structure rule (193) has this restriction.

The above-mentioned differentiation is associated with the insertion of free personal pronouns. As introduced in Chapter 3, free personal pronouns behave like ordinary common phrases, especially in terms of their syntactic positions. The occurrence of free personal pronouns aims to disambiguate the confusion caused by the several co-existing discourse topics; meanwhile, they are switch topics (SWTOPs), supplanting the previous discourse topic in the information flow and foregrounding themselves as the discourse topic in the current information flow. This is manifested in the following sequential sentences excerpted from a ghost story.
The sentences in (195) ends with -sami’ ‘we’, and its referents include both the addressee’s husband and the addressee herself. In order to switch the attention from “us” (i.e. the addressee and her husband) to the addressee’s husband alone, the addressee made use of the free personal pronoun hiya’ ‘he’ as the grammatical subject in (196). This topic lasts to the second sentence in (196), and the null symbol Ø marks both the coreferential empty subject and the continuing topic of the second sentence. The addressee’s husband serving as the discourse topic does not last long since it is soon supplanted by the free personal pronoun kun ‘I’ in (197). Therefore, we propose that free personal pronouns are lexically marked with the discourse feature [+PROM] when they are grammatical subjects, which is able to override the default requirement @DEFAULT (PROM, −) in the phrase structure rule (192) but runs afoul of the mapping constraint with an obligatorily fixed value (↓, PROM) =c − in the phrase structure rule (193).

6.3.2 Object

As for objects in AV clauses, the question-answer tests in Section 6.2.1 show that there is no restriction as to whether the information they convey is given or not since they can accommodate the new information replied in the answer. On the other hand,
the discourse evidence of information chaining in Chapter 5 shows that it is less likely for objects to initiate a new-given information chaining. The information first introduced by grammatical objects can be repeated as an object in the immediately adjacent clause, but it cannot reoccur as a subject unless it undergoes topicalization, a syntactic strategy of foregrounding information. Reversely, given information introduced by objects is not restricted by grammatical functions. The diversity of information is also reflected in statistical analyses. The quantitative studies on TP and RD in Chapter 5 show 57% of objects convey given information while 43% of them new information and that around 41.9% of objects convey important information while 58.1% of them unimportant information. In addition, newness and unimportance are statistically significant respectively. After considering RD and TP at the same time, then we can get the outcome in Table 6-3.

<table>
<thead>
<tr>
<th>Table 6-3 Discourse Features of Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>[GIV] [ABT] AV(%)</td>
</tr>
<tr>
<td>+  +  52 (29.1)</td>
</tr>
<tr>
<td>+  −  50 (27.9)</td>
</tr>
<tr>
<td>−  +  23 (12.8)</td>
</tr>
<tr>
<td>−  −  54 (29.1)</td>
</tr>
<tr>
<td>Total 179</td>
</tr>
</tbody>
</table>

It is demonstrated in Table 6-3 that the distribution is not a gradience like that of subjects; rather, there are three categories of almost equal size, [+GIV, +ABT], [+GIV, −ABT] and [−GIV, −ABT]. Compared with them, the category [−GIV, +ABT] is of smaller size. The outcome of the quantitative evidence dovetails that of the question-answer tests. They both lead to one generalization: the information conveyed by objects is diverse.
Let’s take a close look at the four categories in Table 6-3. The tokens of the first category [+GIV, +ABT] are CTTOPs in i-structure because their information is shared between interlocutors and most of important, as demonstrated in Chapter 5, the given information conveyed by grammatical objects is [+ABT] in nature, being able to connect two clauses easily. The second category [+GIV, −ABT] almost has as many tokens as the first one. It means that grammatical objects also convey background information (BKINF) in i-structure. Unlike what is observed in grammatical subjects, BKINF conveyed by objects is not a non-prototypical backup of CTTOP via the application of \( \theta \); instead, it is a mapping type independent from CTTOP. The one-to-many mapping between f-structure and i-structure here belongs to the independent type, with type variations (Cf. Chapter 4). Like the present thesis, S. Huang & Tanangkingsing (2011) analyzed texts from different sources with Givón’s statistical method as well, and they still found the portion of grammatical object (treated as “extended object” in their paper) conveying important information was up to 50%. Although they did not take a further step to cross-compare [±GIV] and [±ABT] as the present thesis does, their outcome not only buttresses the analysis here that BKINF and CTTOP are different mapping types from object to i-structure, but also proves that text difference plays no role in the present research.

As for the third category in Table 6-3, [−GIV, +ABT], it contains fewer tokens, half smaller than other three categories, which makes its existence unusual. Recall that new information introduced by grammatical objects can only be repeated as an object in the subsequent clause and that the syntactic operation topicalization will be required if it is repeated as a subject or as an oblique in the subsequent clause. Put differently, new information conveyed by objects has to resort to either a discourse strategy like repetition or a syntactic strategy like topicalization so as to get foregrounded in the information flow. It is not important when they are introduced. Therefore, it is proposed that [−GIV, +ABT] does not really enter the correspondences between f-structure and
i-structure. That is to say, \([-\text{GIV}, +\text{ABT}]\) is a syntactically derived variant of the fourth category \([-\text{GIV}, -\text{ABT}]\). The third category does not exist in any on-going discourse, and the phenomenon of their existence is formed by the subsequent discourse.

In contrast, the fourth category \([-\text{GIV}, -\text{ABT}]\) exists independent from the first two. The new unimportant information introduced by objects is not shared between the addressee even mentally or discourse-pragmatically, as illustrated in (198).

\[(198)\] m-luw sa awke nanu’ la \textbf{rgyax} Ø…
AV-go.along that Par Par Part ridge Ø
‘(They) moved along that, m well, ridge…’ <N>

This clause is excerpted from the story about a family’s migration, and its object \textit{rgyax} ‘ridge’ is not only introduced to the story for the first time but also mentally unknown to the addressee, since the addressee cannot predict what kind of route the addresser’s ancestors followed. This information is no longer mentioned in the subsequent discourse. There is no doubt that the fourth category is \textit{CPINF} in i-structure.

Through the above discussion, it is easy to identify the discourse functions to which grammatical objects corresponds: \textit{CTTOP}, \textit{BKINF} and \textit{CPINF}. The phrase structure rule in (199) shows that the complement of an AV verb in c-structure corresponds to a grammatical object in f-structure, as annotated in \((\uparrow\text{OBJ}) = \downarrow\).

\[(199)\] VP \(\rightarrow\) V (DP) \((\downarrow\text{VOICE}) = \text{AV}\) \((\uparrow\text{OBJ}) = \downarrow\)
\[
\begin{align*}
\begin{cases}
\downarrow(\uparrow\text{CTTOP}) = \downarrow \\
\downarrow(\uparrow\text{BKINF}) = \downarrow \\
\uparrow(\uparrow\text{CPINF}) = \downarrow
\end{cases}
\end{align*}
\]
\((\downarrow\text{GIV}) = - \rightarrow (\downarrow\text{ABT}) = \text{c -} @\text{DEFAULT (PROM, -)}\)
In addition, it also corresponds to either of CTTOP, BKINF and CPINF in i-structure, as shown in the annotation \( \{ (\uparrow_i, \text{CTTOP}) = \downarrow, (\uparrow_i, \text{BKINF}) = \downarrow, (\uparrow_i, \text{CPINF}) = \downarrow \} \). They are preceded by their respective symbols of proportion, which are marked by ♥ and ♦. The former means that the probability of the occurrence of CTTOP or BKINF is within the range of 11-30%, while the latter means that the probability of the occurrence of CPINF is within the range of 31-50% (Cf. Chapter 4). There is no restriction on givenness and aboutness so their values are unspecified. However, once the information is new, it has to be unimportant as well, as marked in \( (\downarrow_i, \text{GIV}) = \rightarrow, (\downarrow_i, \text{ABT}) = c \rightarrow \). Moreover, this position (the complement of VP) does not highlight any information unless it is inserted with interrogative words (Cf. Chapter 3), which are lexically marked with [+PROM]. The fixed value in the lexical entry is able to override the constraint annotated to the phrase structure rule in (199), as required by the template @DEFAULT (PROM, −).

### 6.3.3 Oblique

The question-answer tests in Section 6.2.1 show that obliques in NAV clauses are not restricted by the givenness of information, allowing both given and new information. In addition, it is manifest in Chapter 5 that they can initiate given-new information chaining easily no matter what grammatical function the tails are. Nevertheless, the only attested case of its new-given information chaining has to employ topicalization to connect two sentences, indicating that the new information obliques convey is not important to the current information flow.

As we turn to quantitative evidence, we can easily detect that obliques exhibit a strong tendency towards CTTOP. Over 90% of obliques convey given information and more than 70% of obliques convey important information. The outcome of statistical analysis is reorganized in Table 6-4 by making a cross-comparison between [+GIV] and [+ABT]. The category [+GIV, +ABT] is the largest one, occupying 74.5%, while the other three categories, [+GIV, −ABT], [−GIV, +ABT] and [−GIV, −ABT], are rather minor,
occupying only 17.4%, 1.7% and 6.4% respectively.

Table 6-4 *Discourse Features of Obliques*

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>NAV(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>175 (74.5)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>41 (17.4)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>4 (1.7)</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>15 (6.4)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>235</strong></td>
</tr>
</tbody>
</table>

What is more intriguing, this distribution is like what is observed on the subjects of AV clauses, forming gradience. Does the correspondence from oblique to CTTOP also belong to the prototypical type: that is, there is only one mapping type but with token variations? If only quantitative evidence is taken into consideration, the answer would be yes. However, the question-answer tests do not support this analysis since obliques are different from the subjects of AV clauses in that the former are not required to convey given information. In other words, the two smallest categories are not the backups of the prototypical one [+GIV, +ABT]; instead, they are mapping types independent from the major one, but with much smaller scale. The gradience that the four categories form is merely a coincidence. In addition, the third category [−GIV, +ABT] actually does not exist because the occurrence of [+ABT] is due to the use of either the syntactic strategy topicalization or the discourse strategy repetition (Cf. Chapter 5), like what have been observed on grammatical objects.

As for the fourth category, new and unimportant information does exist, as illustrated in (200). The participant *laqi-mu* ‘my child’ is introduced by the oblique into the information flow for the first time, but it is not repeated anymore in the subsequent discourse.
(200) i... qutux hi’ la-ga, qnyat-an laqi=mu’.
Par one body Par-Top raise-LV child=1SGen
‘Well, (speaking of) one (of them), it is raised by my child.’ <N>

The phrase structure rules in (201) and (202) are annotated with the correspondences between c-structure and f/i-structure, the former of which applies to nominals while the latter of which applies to pronominal clitics.

(201) I’  \rightarrow  I'  

(DP)  
(↓ VOICE) = NAV  
(↑ OBL) = ↓  
(↓ CASE) = OBL  
{◆(↑₁ CTTOP) = ↓}  
{♥(↑₁ BKINF) = ↓}  
{♥(↑₁ CPINF) = ↓}  

(↓₁ GIV) = − → (↓₁ ABT) =c−  
@DEFAULT (PROM, −)

(202) I’  \rightarrow  I  

(CL)  
(↓ VOICE) = NAV  
(↑ OBL) = ↓  
(↓ CASE) = OBL  
(↓ PRED) = ‘PRO’  
{◆(↑₁ CTTOP) = ↓}  
{♥(↑₁ BKINF) = ↓}  
{♠(↑₁ CPINF) = ↓}  

(↓₁ GIV) = − → (↓₁ ABT) =c−  
(↓₁ PROM) =c−

The difference between the above phrase structure rules lies in the existence of (↓ PRED) = ‘PRO’, which means that the DP must be a pronoun. The DP adjoined to I’ or cliticized to I in c-structure corresponds to an oblique in f-structure, as marked in (↑ OBL) = ↓, and its case marking is oblique as shown in (↓ CASE) = OBL. The discussion in previous paragraphs shows that obliques in f-structure correspond to CTTOP, BKINF and CPINF in
i-structure. They belong to three different mapping types, further leading the one-to-many mapping from f-structure to i-structure. The distinction between non-clitics and clitics does not change what discourse functions in i-structure obliques correspond to but only slightly change the percentage of each discourse function. The most obvious difference is that non-clitics has a lower percentage of the first category [+GIV, +ABT] and a higher percentage of the four category [−GIV, −ABT], as illustrated in Table 6-5.

Table 6-5 Discourse Features of Obliques (Non-clitics vs. Clitics)

<table>
<thead>
<tr>
<th></th>
<th>[GIV]</th>
<th>[ABT]</th>
<th>NAV(%)</th>
<th>NAV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Non-clitics</td>
<td>Clitics</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>51 (65.4)</td>
<td>124 (79.0)</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>14 (17.9)</td>
<td>27 (17.2)</td>
<td></td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>1 (1.3)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>12 (15.4)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>78</strong></td>
<td><strong>157</strong></td>
<td></td>
</tr>
</tbody>
</table>

The mapping between c-structure and i-structure is marked in the annotation \( \{ \uparrow CTTOP = \downarrow, \uparrow BKINF = \downarrow, \uparrow CPINF = \downarrow \} \), with the precedence of their respective symbols of proportion, which are marked by ♠, ♣, ♥ and ●. The first one indicates the probability within the range of 71-90%, the second one indicates the probability within the range of 51-70%, the third one the probability within the range of 11-30%, and the fourth one the probability within the range of 1-10% (Cf. Chapter 4). There is no restriction on givenness and aboutness so they are annotated here. If the information is new, it has to be unimportant, as marked in \( \downarrow GIV = \rightarrow (\downarrow ABT) = \ast \rightarrow \). In addition, since interrogative words are permitted being inserted into the DP adjoined to \( \Gamma \) and they are lexically marked with [+PROM], it is proposed that the value of
prominence in (201) can be overridden, not being obligatorily fixed, as shown in the template @DEFAULT (PROM, −).

6.4 Explanations for Empirical Puzzles

Based on the phrase structure rules (192), (193) and (199) in Section 6.3, the unmarked interactions between c-structure, f-structure and i-structure in AV clauses are presented in (203) and (204). The former represents the mapping with non-clitics, and the latter represents the mapping with a clitic subject. On the other hand, after assembling the phrase structure rules (192), (193), (201) and (202) in Section 6.3, we can get the unmarked interactions of NAV clauses in (205), (206), (207) and (208). The mapping in (205) concerns two nominal arguments while the others involve clitic arguments, with a clitic oblique in (206), a clitic subject in (207) and two clitic arguments in (208). Their main differences mainly lie in the configuration of phrase structures in c-structure. For the sake of clarity, the mapping between c-structure and f-structure is indicated by green lines, the mapping between f-structure and i-structure is indicated by purple lines, and the mapping between c-structure and i-structure is indicated by pink lines. The independent type of one-to-many mapping is substantiated with one mapping line as well as bracketed discourse functions. The proportion of each discourse function in one-to-many mapping is labeled with the symbols of proportion. The non-prototypical back-up candidates for the correspondence between CTTOP in i-structure and SUBJ in f-structure are vertically placed below CTTOP and they undergo the conversion function $ϙ$, illustrated with a blue line, to enter the general mapping between i-structure and other representations via communicative cooperation and cognitive ability.

These correspondences formalized and visualized in the model of the present thesis benefit explaining some empirical puzzles in typological linguistics. They will be discussed one by one in the rest of this section.
(203) Mapping in AV clauses with non-clitics
(204) Mapping in AV clauses with a clitic subject

c-structure

f-structure

SUBJ [ SUBJ ]
OBJ [ OBJ ]

VOICE AV

i-structure

CTTOP [ GIV + ]
[ ABT + ]
[ PROM - ]

♥CTTOP [ GIV + ]
[ ABT + ]
[ PROM - ]

♥BKINF [ GIV + ]
[ ABT - ]
[ PROM - ]

♦CPINF [ GIV - ]
[ ABT - ]
[ PROM - ]

BKINF [ GIV + ]
[ ABT - ]
[ PROM - ]
(205) Mapping in NAV clauses with non-clitics

c-structure

f-structure

SUBJ [ ]
OBL [ ]
VOICE NAV

i-structure

\[ \begin{align*}
\text{\textbullet CTTOP} & : \begin{cases}
GIV & + \\
ABT & + \\
PROM & - \\
GIV & +
\end{cases} \\
\text{\textbullet BKINF} & : \begin{cases}
GIV & - \\
ABT & - \\
PROM & - \\
GIV & -
\end{cases} \\
\text{\textbullet CPINF} & : \begin{cases}
GIV & - \\
ABT & - \\
PROM & -
\end{cases}
\end{align*} \]
(206) Mapping in NAV clauses with a clitic oblique

c-structure

\[
\begin{array}{c}
\text{IP} \\
\text{I'} \\
\text{L-DP} \\
\end{array}
\]

f-structure

\[
\begin{array}{c}
\text{SUBJ} \quad [ ] \\
\text{OBL} \quad [ ] \\
\text{VOICE} \quad \text{NAV} \\
\end{array}
\]

i-structure

\[
\begin{array}{c}
\text{♣CTTOP} \\
\text{♥BKINF} \\
\text{●CPINF} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{GIV} + \} \\
\{ \text{ABT} + \} \\
\{ \text{PROM} - \} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{GIV} + \} \\
\{ \text{ABT} - \} \\
\{ \text{PROM} - \} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{CTTOP} \} \\
\{ \text{PSFOC} \} \\
\{ \text{CPINF} \} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{GIV} + \} \\
\{ \text{ABT} - \} \\
\{ \text{PROM} - \} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{GIV} - \} \\
\{ \text{ABT} + \} \\
\{ \text{PROM} - \} \\
\end{array}
\]

\[
\begin{array}{c}
\{ \text{GIV} - \} \\
\{ \text{ABT} - \} \\
\{ \text{PROM} - \} \\
\end{array}
\]
(207) Mapping in NAV clauses with a clitic subject

c-structure

f-structure

SUBJ [ ]
OBL [ ]
VOICE NAV

i-structure

CTTOP [GIV +]
[ABT +]
PROM -]
Q

BKINF [GIV +]
[ABT -]
PROM -]

❣️CTTOP [GIV +]
[ABT +]
PROM -]
 szczególny

❣️BKINF [GIV +]
[ABT -]
PROM -]

❣️CPINF [GIV -]
[ABT -]
PROM -]
(208) Mapping in NAV clauses with a clitic subject and a clitic oblique

c-structure

f-structure

i-structure

\[
\begin{align*}
\text{c-structure} & : \quad \text{IP} \\
& \quad \text{I} \\
& \quad \text{I-DP-DP} \\
\text{f-structure} & : \quad \lbrack \text{SUBJ} \rbrack \\
& \quad \lbrack \text{OBL} \rbrack \\
& \quad \text{VOICE NAV} \\
\text{i-structure} & : \quad [\{ \begin{array}{c} \text{\textbullet CTTOP} \\
& \quad \text{\textbullet BKNF} \\
& \quad \text{\textbullet CINF} \\
\end{array} \}
\begin{array}{c}
\text{GIV} \\
\text{ABT} \\
\text{PROM} \\
\end{array}
\begin{array}{c}
+ \\
+ \\
- \\
\end{array}
\begin{array}{c}
\text{GIV} \\
\text{ABT} \\
\text{PROM} \\
\end{array}
\begin{array}{c}
- \\
- \\
- \\
\end{array}
\begin{array}{c}
\text{CTTOP} \\
\text{ABT} \\
\text{PROM} \\
\end{array}
\begin{array}{c}
+ \\
+ \\
- \\
\end{array}
\begin{array}{c}
\text{BKNF} \\
\text{ABT} \\
\text{PROM} \\
\end{array}
\begin{array}{c}
+ \\
- \\
- \\
\end{array}
\begin{array}{c}
\text{IP} \\
I' \\
\text{I-DP-DP} \\
\end{array}]
\end{align*}
\]
6.4.1 F-Structural vs. I-Structural

According to Mithun (1987), languages in the world are divided into two typological types in terms of their word order, syntactically based languages and pragmatically based languages. The word order in the syntactically based languages is mainly determined by grammatical functions, whereas the word order in the pragmatically based languages is mainly determined by pragmatic considerations. On the assumption that c-structure is the output of the theoretical model, as advocated in LFG-OT, syntactically based languages and pragmatically based languages may be termed as f-structural languages and i-structural languages respectively.91 What Squliq Atayal is in this typological classification is displayed in the inter-modal mapping of clauses with non-clitics, as presented in (203) and (205). On the one hand, f-structure is a crucial determinant of c-structure because there is one-to-one mapping both between the outer DP in c-structure and SUBJ in f-structure and between the inner DP in c-structure and OBJ/OBL in f-structure. On the other hand, i-structure is a partial determinant of c-structure because there is one-to-one mapping between the outer DP in c-structure and CTTOP in i-structure as well. That is to say, Squliq Atayal is not a radical f-structural language but just a quasi-f-structural language since part of its c-structure, particularly the outer DP is determined by i-structure as well. This theoretical model not only dovetails the outcome of the analysis of multinominal logistic regression in Appendix 1 but also correctly visualizes it.

91 The distinction between f-structural languages and i-structural languages may not be able to exhaust typological diversity. From the perspective of linguistic typology, there exist a-structural languages, whose word order is mainly determined by thematic functions. For instance, the unmarked word order in Amis and Isbukun Bunun (Austronesian languages) is Verb-Actor-Patient (Huang & Shi 2016, Wu 2000, 2006). In addition, some languages like Achang and Gyalrong (Tibeto-Burmese languages) mark nouns by their thematic roles, not by their grammatical functions (Lin 1993, Shi 2009). They are called languages of “semantically based marking” in Dixon (1994). It is intriguing whether their word order is like their morphological marking in being mainly determined by thematic functions though morphological marking and word order are not mutually dependent. Even though there may be no radical a-structural language, due to the collaborations between modules, the possibility of the existence of quasi-a-structural languages is still high.
In addition, it is rational to claim that Squilq Atayal is a language with both topic-prominence and subject-prominence since the relationship between the outer DP and the rest of the clause in c-structure reflects not only the subject-predicate relation but the topic-comment relation as well. This reminds us that typological classifications, such as f-structural languages vs. i-structural languages, or subject-prominent languages vs. topic-prominent languages, always have the deficiency of discreteness, that is, facing the challenge of prototype, gradience, and continuum (Aarts 2007, Taylor 1995). That is why thetic judgments fail to make a clear-cut distinction between subject-prominent languages and topic-prominent languages (Cf. Section 6.2.2). Even the proposers of the above typological classifications never repudiate the existence of gray area. The correspondences shown in (203) and (205) not only present this gray area but also pinpoint its specific position: that is, the outer DP, which has one-to-one mapping with both f-structure and i-structure.

6.4.2 The Essence of Multiple Voices

6.4.2.1 Information-Salience

According to Klaiman (1991), grammatical voice is a category of mixed concepts, and there are three types of voice systems in linguistic typology. One is the role-remapping voice system, where the alternation of verbal morphology signals the remapping between grammatical functions and thematic functions, like the active-passive alternation in English, another is the alternation in the subject’s participant status, where the alternation of verbal morphology reflects the affectedness of logic subject (i.e. actor), like the active-middle alternation in Tamil, and the other is the information-salience voice system, where the alternation of verbal morphology encodes “nominals’ relative centrality or noncentrality to the discourse’s informational objectives (Klaiman 1991:228)”, like the multiple voice system in Philippine languages.
Although a-structure is not displayed in the correspondences from (203) to (208), the one-to-one mapping between SUBJ and CTTOP unveils an intriguing aspect of the voice system in Squiliq Atayal: that is, no matter what thematic function in a-structure is chosen to correspond to SUBJ via the voice system, it also corresponds to CTTOP. In other words, what the voice system in this language selects is not only SUBJ, the only syntactically active grammatical function, but also CTTOP, the most important information shared between interlocutors. Shibatani (1988a, b) made a similar proposal on Philippine languages that the voice system in Philippine languages selects both subject and topic. Since Squiliq Atayal and Philippine languages are genetically related, all in the Austronesian family, it is not unexpected that the voice system in Squiliq Atayal belongs to the information-salience voice system as well. However, it was not settled in previous studies exactly what the above-mentioned topic is in Philippine languages. They captured the core of these languages but used the poorly-defined term “topic”, such as Starosta, Pawley & Reid (1982). The identification of what this topic is (in fact, it is CTTOP) is one of the contributions of the present thesis.92

However, what we have to emphasize is that even the information-salience voice system still involves the mechanism of the role-remapping voice system since the outer DP in (203) and (205) is both SUBJ and CTTOP. The role-remapping voice system and the information-salience voice system are not mutually exclusive. Squiliq Atayal has both of them. If we only look at the mapping between f-structure and a-structure, it is of the role-remapping voice system, whereas if we only look at the mapping between

---

92 Although Givón (1984, 1990) and Mithun (1987) already provided the insight that the grammatical subject in the word order VOS conveys continuing information, they did not clearly define and classify different types of information statuses/discourse functions in their works. As mentioned for several times in the present thesis, although we do believe that continuum/gradience is the norm of linguistic reality, classification in traditional analysis still has its value in helping us differentiating prototypical tokens from non-prototypical ones. Overemphasis on continuum will cause the messy mix of distinct categories and may hinder future research. For instance, Shibatani (1988b) proposed that there was a continuum between topic and subject to account for the dual role of grammatical subject in Philippine languages. Although his analysis was insightful, his comparisons between Japanese-type and Philippine-type topicalization in the endnote unnecessarily lumped together voice alternation and topicalization. In reality, topicalization in Formosan languages and Philippine languages is usually manifested by its marked word order whereas voice alternation does not affect unmarked word order.
i-structure and a-structure, it is of the information-salience system. Even in English, which has the role-remapping voice system, voice alternation still has something to do with i-structure, not just being a matter between f-structure and a-structure. For instance, subject of long passives must not be less familiar in the discourse than oblique, which is inclined to introduce new information (Biber et al. 1999, Ward, Birner & Huddleston 2002). The disparity between Squliq Atayal and English is that although they both comply with the given-new discourse principle (or the Given-First Constraint in Chapter 5) to certain extent, the former requires the one-to-one consistent mapping between SUBJ and CTTOP while the latter does not have this requirement, permitting SUBJ to convey new information (Lambrecht 1994). Without considering the inherent properties of tense and aspect, our parallel approach manifests that the voice system in Squliq Atayal is verbal morphology encoding the alternation of a-structure’s mapping to f-structure and i-structure. Once we realize that grammatical voice is merely one of the ways to encode correspondences between representations, it is not difficult to comprehend why languages without grammatical voice exist, such as Warlpiri (Nash 1986, Simpson 1991), since the mechanisms behind grammatical voice can be encoded in other ways.

Recall that it is problematic for Riesberg’s (2014) lexically determined mapping approach to predict the mapping between f-structure and a-structure because one voice marking may select several possible thematic functions as SUBJ (Cf. Chapter 3). Take (209) for instance. The main LV verb gyahan ‘open’ selects either PAT or LOC as SUBJ, but the lexical determined mapping in symmetrical languages only makes the latter possible. The former mapping is possible under the passivization constraint of asymmetrical languages but Squliq Atayal is not an asymmetrical language.
Riesberg’s (2014) lexically determined mapping approach is on the right track because symmetrical languages treat multiple voices equal by marking all voices. Its weakness comes from the fact that the whole mechanism is built on the role-remapping voice system, with a narrow scope on the mapping between a-structure and f-structure, which is not precise to select SUBJ from the inventory containing PAT and LOC. If a broader view is taken, with the mapping with i-structure (i.e. the essence of the information-salience voice system) being considered, the one-to-one mapping between SUBJ and CTTOP will filter out inappropriate candidates. For instance, when the lexical property of LV marking -an maps PAT to SUBJ, the required mapping between SUBJ and CTTOP will be satisfied if PAT maps to CTTOP in context, as illustrated in (210); otherwise, the whole mapping will crash as illustrated in (211).
In the same vein, when the lexical property of LV marking -an maps LOC to SUBJ, the required mapping between SUBJ and CTTOP will be satisfied if LOC maps to CTTOP in context, as illustrated in (212); otherwise, this required mapping will not be satisfied, as illustrated in (213).

Even when both PAT and LOC map to CTTOP, there will not be two subjects because the
SUBJ mapping between a-structure and f-structure is only applied for once.

Why are there four different voices in Squliq Atayal? The answer to this question is closely related to the dual nature of multiple voices. Multiple voices admit of more flexibility in selecting the salient continuing topic in discourse (CTTOP in i-structure) from diverse thematic functions.

6.4.2.2 Comparisons with English

The previous section has presented that information-salience is the fundamental difference between voices in Squliq Atayal and in English, in terms of the mapping between a-structure, f-structure and i-structure. However, this section will make more comparisons between them by examining how actor and non-actor correspond to c-structure, f-structure and i-structure respectively in a transitive clause.

First of all, when it comes to c-structure, AV clauses in Squliq Atayal resembles active clauses in English while NA V clauses in Squliq Atayal resembles passive clauses in English, as shown in Table 6-6:

<table>
<thead>
<tr>
<th>actor</th>
<th>non-actor</th>
<th>Squiliq Atayal</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>specifier</td>
<td>complement</td>
<td>AV</td>
</tr>
<tr>
<td>b.</td>
<td>adjunct</td>
<td>specifier</td>
<td>NAV</td>
</tr>
</tbody>
</table>

Regardless of the difference in the word order, both AV clauses in Squliq Atayal and active clauses in English have the actor as a specifier and the non-actor as a complement in c-structure. In the same vein, both NAV clauses in Squliq Atayal and passive clauses in English have the actor as an adjunct and the non-actor as a specifier in c-structure. The evidence of Squliq Atayal has been thoroughly discussed in Chapter 3, not repeated here again.
The second comparison between Squiliq Atayal and English concerns the thematic functions actor and non-actor convey in f-structure, as shown in Table 6-7.

Table 6-7 Comparison II: F-Structure

<table>
<thead>
<tr>
<th>actor</th>
<th>non-actor</th>
<th>Squiliq Atayal</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>SUBJ</td>
<td>OBJ</td>
<td>AV</td>
</tr>
<tr>
<td>b.</td>
<td>OBL</td>
<td>SUBJ</td>
<td>NAV</td>
</tr>
</tbody>
</table>

The issue on the identification of grammatical functions has been discussed in Chapter 3. Briefly speaking, the constituency between the non-actor and the verb in AV clauses proves that the grammatical function of the non-actor in AV clauses is OBJ. In addition, the existence actorless NAV clauses and the communtability between the actor and temporal adverbials in NAV clauses proves that the grammatical function of the actor in NAV clauses is OBL. The phenomenon that the actor conveys SUBJ and the non-actor conveys OBJ in f-structure is observed in both of AV clauses in Squiliq Atayal and active clauses in English. Likewise, the phenomenon that the actor conveys OBL and the non-actor conveys SUBJ is observed in both of NAV clauses in Squiliq Atayal and passive clauses in English. Once again, AV and NAV in Squiliq Atayal look equivalent to active voice and passive voice in English respectively. The only difference from English is the split-subjecthood phenomena: that is, the actor in NAV clauses has the anaphoric prominence and can serve as the addressee of imperatives. They are closely related with information structure, not purely syntactic, and this issue will be elaborated in Chapter 9.

Although the above two comparisons between Squiliq Atayal and English may be difficult to be accepted by some Austronesian linguists who tend to emphasize the difference between Austronesian languages and Indo-European languages, there is no denying that at least in c-structure and f-structure, the traditional core of syntax,
similarities do exist to some extent between these two languages. After all, both Squliq Atayal and English are configurational languages, in which configuration associated with both c-structure and f-structure is the best means to distinguish difference voices. In addition, it is suggested that the notions of promotion and demotion should not be applied to the relationship between actor and non-actor in voice systems since they always imply the misconception that one voice is syntactically derived or transformed from the other voice(s). As a matter of fact, voice alternation is not a transformational rule applying to all words but a set of lexical mapping constraints with some obvious lexical exceptions (Cf. Section 3.3.4). For instance, in English, some transitive verbs, like have, cost, fit, resemble, befall, etc., cannot be passivized while some passivized verbs, like be born, be reputed, etc., have almost lost their active counterparts (Huddleston 2002, Quirk et al. 1985, among others). By incorporating the CG concept into LFG, we may as well claim that AV and NA V are mutually independent constructions in Squliq Atayal and so are active and passive in English. A construction is a composite of parallel representations (like a-structure, c-structure, f-structure and i-structure) with the establishment of mutual association through correspondences. That is to say, when it comes to the relationship between the non-actor in AV clauses and its counterpart in NA V clauses, it is more appropriately to regard it as the non-actor’s mapping to different grammatical functions in different constructions than to attach it with the unnecessary directional notion of the non-actor’s promotion from non-subject to subject. The promotion/demotion in syntax does not equal the promotion/demotion in

---

93 As for the active counterparts of be born and be reputed, their latest examples in Oxford English Dictionary Online (December 2016) are cited from texts written in the 19th century and the 18th century respectively:

(i) 1855  C. Kingsley Heroes (1868) i. 2  Your daughter Danae shall bear a son.
(ii) 1790  E. Burke Refl. Revol. in France 184  Is it then impossible that a man may be found… who may repute that nation to be destitute of all wisdom and of all virtue, which…thought proper to commit a thousand crimes.

94 The asymmetry in voice alternation is commonly observed in languages. According to Zhang (2016), in French, some transitive verbs, like avoir ‘have’ and comporter ‘contain’, tend not to be passivized while some passivized expressions, like être bien vu ‘be respected’ and être mal vu ‘be despised’, lack the active counterparts. The asymmetry in voice alternation is part of the evidence which Starosta (2002) bases to claim that the voice marking in Austronesian languages is derivational in nature.
discourse-pragmatics, and vice versa. Eventually what promotion/demotion is will become obscure—different studies may have different bases for promotion/demotion—, unable to clarify the essence of voice system.

The third comparison between the voice system in Squiliq Atayal and in English focuses on i-structure alone, particularly the issue on the coexistence of given and new information in a proposition. When two pieces of information coexist in a proposition in Squiliq Atayal, there are three possible patterns: one piece of given information and one piece of new information, two pieces of given information, and two pieces of new information. Based on the data of the present thesis, the proportion of each coexistence pattern in AV clauses and NAV clauses is shown in Table 6-8.

<table>
<thead>
<tr>
<th>Coexistence Patterns</th>
<th>AV (%)</th>
<th>NAV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [+GIV] [−GIV]</td>
<td>79 (44.1)</td>
<td>43 (19.6)</td>
</tr>
<tr>
<td>b. [+GIV] [+GIV]</td>
<td>96 (53.6)</td>
<td>173 (79.0)</td>
</tr>
<tr>
<td>c. [−GIV] [−GIV]</td>
<td>4 (2.2)</td>
<td>3 (1.4)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>179</td>
<td>219</td>
</tr>
</tbody>
</table>

The major coexistence pattern is the b-pattern, i.e. a proposition with two pieces of given information. This fact is of no surprise because the same characters or items are always recurrent in one text. The proportion of the c-pattern is too low, so it is not discussed here. What is really intriguing is the a-pattern, i.e. a proposition with one piece of given information and one piece of new information, because these two pieces of information are opposite to each other in terms of givenness.

The coexistence of given and new information can be further divided into two patterns, as shown in Table 6-9.
When actor conveys given information and non-actor conveys new information, it is the first pattern, the a-pattern in the above table. Conversely, when actor conveys new information and non-actor conveys given information, it is the second pattern, the b-pattern in the above table. It is demonstrated in Table 6-9 that the a-pattern tends to occur in AV clauses (74.7% in AV clauses and 25.3% in NAV clauses) whereas the b-pattern tends to occur in NAV clauses (21.7% in AV clauses and 78.3% in NAV clauses). That is to say, the actor in AV clauses usually conveys given information while the non-actor in NAV clauses usually conveys new information. This distribution is accordance with the fact that multiple voices in Squiq Atayal belong to the information-salience voice system.

In addition, the same statistical data can be analyzed from another angle. This is shown in the following table:

<table>
<thead>
<tr>
<th>actor</th>
<th>non-actor</th>
<th>AV (%)</th>
<th>NAV (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[+GIV]</td>
<td>[-GIV]</td>
<td>74 (74.7)</td>
<td>25 (25.3)</td>
</tr>
<tr>
<td>b.</td>
<td>[-GIV]</td>
<td>[+GIV]</td>
<td>5 (21.7)</td>
<td>18 (78.3)</td>
</tr>
</tbody>
</table>

Table 6-10 Coexistence of Given and New Information (II)

<table>
<thead>
<tr>
<th>actor</th>
<th>non-actor</th>
<th>AV (%)</th>
<th>NAV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[+GIV]</td>
<td>[-GIV]</td>
<td>74 (93.7)</td>
</tr>
<tr>
<td>b.</td>
<td>[-GIV]</td>
<td>[+GIV]</td>
<td>5 (6.3)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>79</td>
</tr>
</tbody>
</table>

The above table takes the voice-internal perspective. In AV clauses, the a-pattern is the absolute majority, up to 93.7%. That is to say, the information-structural motive of AV clauses is to make the actor convey given information and the non-actor convey new
information. In NAV clauses, both of the a-pattern and the b-pattern play an important role, which shows that NAV clauses accommodate diverse coexistence patterns. Similar observation has been presented in Chapter 5.

Based on the above observation, the third comparison between Squiliq Atayal and English is shown in the following table:

Table 6-11 *Comparison III: I-Structure*

<table>
<thead>
<tr>
<th>actor</th>
<th>non-actor</th>
<th>Squiliq Atayal</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[+GIV]</td>
<td>[−GIV]</td>
<td>AV</td>
</tr>
<tr>
<td>b.</td>
<td>[−GIV]</td>
<td>[+GIV]</td>
<td>NAV</td>
</tr>
</tbody>
</table>

The a-pattern in the above table, where the actor conveys given information and the non-actor conveys new information, occurs in both of AV clauses and NAV clauses in Squiliq Atayal while it mainly occurs in active clauses in English. In contrast, the b-pattern in the above table, where the actor conveys new information and the non-actor conveys given information, occurs in NAV clauses in Squiliq Atayal and in passive clauses in English. The distribution of these two coexistence patterns is one of the most important differences between Squiliq Atayal and English. Unlike what is shown in the previous two comparisons, in terms of i-structure, particularly the coexistence of given and new information, there is a remarkable mismatch between AV & active and between NAV & passive. AV clauses and NAV clauses in Squiliq Atayal are not thoroughly equivalent to active clauses and passive clauses in English respectively.

Generally speaking, with regard to voice systems, the similarities and differences between Squiliq Atayal and English vary from representation to representation. In c-structure and f-structure, that is, at the level of syntax, these two languages are similar.
to each other. By contrast, in i-structure, that is, at the level of discourse-pragmatics, their difference is much more remarkable. The parallelism of LFG helps pinpoint where the similarities and differences between the voice system in Squiliq Atayal and in English really are.

6.4.3 Explanation of VOS

Why does Squiliq Atayal have Verb-Object/Oblique-Subject (VOS) as its unmarked word order, being one of the minority languages with this characteristic in the world? Information-based accounts usually fail to elaborate the correlation between word orders and information statuses even though they seem to be better suited for accounting for non-conventionalized or non-grammaticalized within-language word-order variation (Hawkins 1994, Song 2012). However, without i-structure, it would be difficult to explain why the word order VOS is rare in linguistic typology since phrase structure rules in the LFG model or the parameter setting in MGG alone has no preference on any particular word order. The correlation between c-/f-structure and i-structure presented so far in the present thesis is definitely neither coincidental nor meaningless.

In the theoretical model presented in the present thesis, i-structure as the representation of discourse-pragmatics is as essential as other representations but its influence on word order varies from language to language, correlated with two mechanisms—correspondences and i-structure constraints. Correspondences have been discussed in Section 6.4.1, where it is shown that Squiliq Atayal is a quasi-f-structural language, with the stronger influence of f-structure and the weaker influence of i-structure both upon c-structure. Look back at (203) and (205). If the conversion function \( \varnothing \) is tentatively put aside, the linear order of discourse functions in i-structure constitutes only three patterns, that is, a., b. and c. in (214) and (215).
(214) I-Structure Constraint Ranking (I): AV Clauses

<table>
<thead>
<tr>
<th></th>
<th>OBJ</th>
<th>SUBJ</th>
<th>Aboutness-Last</th>
<th>Given-First</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>CTTOP</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV +]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT +]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>BKINF</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV +]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT −]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>CPINF</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV −]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT −]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(215) I-Structure Constraint Ranking (II): NAV Clauses

<table>
<thead>
<tr>
<th></th>
<th>OBL</th>
<th>SUBJ</th>
<th>Aboutness-Last</th>
<th>Given-First</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>CTTOP</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV +]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT +]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>BKINF</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV +]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT −]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>CPINF</td>
<td>CTTOP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[GIV −]</td>
<td>[GIV +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ABT −]</td>
<td>[ABT +]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PROM −]</td>
<td>[PROM −]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In terms of the linearity of information functions, CTTOP precedes CTTOP in the a-pattern while BKINF precedes CTTOP in the b-pattern. In these two patterns, given information is followed by given information, which does not violate the requirement of the Given-First Constraint that given information must come first when two pieces of information are adjacent. In addition, these two patterns follow the requirement of the Aboutness-Last Constraint that important information to subsequent discourse must
come last if two pieces of information are adjacent. Their obedience to these two constraints is reflected to the empty blanks under the constraint columns in a-pattern and b-pattern respectively.

Unlike the first two patterns, c-pattern does not fulfill both of these two i-structure constraints. The linear order that CPINF precedes CTTOP violates the Given-First Constraint since the given information does not come first, although this patterns still complies with the Aboutness-Last Constraint. The violation is illustrated in the asterisk marked in the column Given-First of the c-pattern. Therefore, the c-pattern proves that obedience of the Aboutness-Last Constraint is more important than that of the Given-First Constraint in this language. Following the LFG-OT convention, we may claim that Aboutness-Last outranks over Given-First. It is not necessary to adopt the LFG-OT convention since this ranking is interpreted as follows: the linear order of discourse functions in i-structure prefers the coherence with subsequent discourse to the coherence with previous discourse especially on the condition that one of these two types of coherence has to be sacrificed. This ranking in i-structure yields its influence on the word order in c-structure through correspondences between i-structure and c-structure.

As stated in Croft (2003), “[t]he competing motivations are ultimately functional in origin (...), and are present for all speakers of all languages. Competing motivations influence (i.e. constrain) the conventions of a grammar of a particular language, and can alter the conventions in the process of language change (...). But they are not part of the conventions of a particular language at a particular time per se (p. 64).” The i-structural constraints are representations of competing functional motivations, but the above rankings of i-structural constraints are merely an ideal pattern because since the application of the conversion function Q in i-structure brings aberrant cases in naturalistic data. That is why the rankings in (214) and (215) never stand in reality, “not part of the conventions of a particular language,” if all tokens are lumped together.
without differentiating mapping type from mapping token.

If taking into consideration the discourse functions entering the mapping between i-structure and f-structure through $Q$, we will get the constraint rankings in (216) and (217) below.
### (216) I-Structure Constraint Ranking (III): AV Clauses

<table>
<thead>
<tr>
<th></th>
<th>OBJ</th>
<th>SUBJ</th>
<th>Aboutness-Last</th>
<th>Given-First</th>
</tr>
</thead>
<tbody>
<tr>
<td>d.</td>
<td>CTTOP</td>
<td>GIV +</td>
<td>GIV +</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT +</td>
<td>ABT +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>BKINF</td>
<td>GIV +</td>
<td>GIV +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>CPINF</td>
<td>GIV −</td>
<td>GIV +</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>CTTOP</td>
<td>GIV +</td>
<td>GIV −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT +</td>
<td>ABT +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>h.</td>
<td>BKINF</td>
<td>GIV +</td>
<td>GIV −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>i.</td>
<td>CPINF</td>
<td>GIV −</td>
<td>GIV −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>j.</td>
<td>CTTOP</td>
<td>GIV +</td>
<td>GIV −</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT +</td>
<td>ABT −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>k.</td>
<td>BKINF</td>
<td>GIV +</td>
<td>GIV −</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
<tr>
<td>l.</td>
<td>CPINF</td>
<td>GIV −</td>
<td>GIV −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABT −</td>
<td>ABT −</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROM −</td>
<td>PROM −</td>
<td></td>
</tr>
</tbody>
</table>
I-Structure Constraint Ranking (IV): NAV Clauses

<table>
<thead>
<tr>
<th></th>
<th>OBL</th>
<th>SUBJ</th>
<th>Aboutness-Last</th>
<th>Given-First</th>
</tr>
</thead>
<tbody>
<tr>
<td>d.</td>
<td>CTTOP</td>
<td>BKINF</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>BKINF</td>
<td>BKINF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>CPINF</td>
<td>BKINF</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>CTTOP</td>
<td>PSFOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h.</td>
<td>BKINF</td>
<td>PSFOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i.</td>
<td>CPINF</td>
<td>PSFOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j.</td>
<td>CTTOP</td>
<td>CPINF</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>k.</td>
<td>BKINF</td>
<td>CPINF</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>l.</td>
<td>CPINF</td>
<td>CPINF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The patterns from d. to f. have SUBJ corresponding to BKINF, and this causes the violation of **Aboutness-Last** in the d-pattern, which runs afoul of the ideal ranking in (214) and (215). Similarly, the patterns from j. to l. conflict with the ideal ranking either, with the violation of **Aboutness-Last** in the j-pattern. The ideal ranking is obeyed only in the patterns from g. to i., with PSFOC corresponding to SUBJ. These conflicts are not meaningless. What they reflect is a gradually slow change in i-structure: that is, the Given-First Constraint is on the way to rank over the Aboutness-Last. This process may take a long while to accomplish, and it only exists in i-structure, not expanding to the mapping from i-structure to other representations for the moment. Definitely, it is still relevant to the application of the conversion function $Q$. The details of this diachronic process will be thoroughly discussed in Chapter 9.

Generally speaking, the formation of VOS in Squiliq Atayal is closely related to the ideal ranking of i-structural constraints, which favors the Aboutness-Last Constraint that the information connecting with subsequent discourse comes later over the Given-First Constraint that the information connecting with previous discourse comes first. Therefore, the grammatical subject, which is a continuing topic as well, has to come last, following other arguments. Strictly speaking, the ideal ranking only exists in mapping types, and it may have altered the grammar of this language in history but cannot be attested by mapping tokens since the application of discourse-pragmatic forces like the conversion function has always been working in i-structure.

### 6.5 Summary

This chapter employs syntax-oriented tests and the discourse evidence discussed in Chapter 5 to identify what discourse functions the core arguments (subject, object and oblique) convey in i-structure. Subject only corresponds to CTTOP in i-structure and the superficial one-to-many mapping phenomenon from the quantitative outcome is attributed to the application of the conversion function $Q$, which originates from the
cooperation between interlocutors in the process of communication. When it is applied, discourse functions like BKINF, PSFOC and CPINF are able to enter the default mapping between i-structure and f-structure, a substantiation of the prototypical type of one-to-many mapping. On the other hand, object and oblique both correspond to CTTOP, BKINF and CPINF in i-structure. The correspondences belong to different mapping types, a substantiation of the independent type of one-to-many mapping.

The general mapping between representations shows that although there is a more consistent one-to-one mapping between f-structure and c-structure, i-structure still plays an essential role particularly for subject; in other words, Squiliq Atayal is a quasi-f-structural language. The dual role of grammatical subject in (SUBJ in f-structure and CTTOP in i-structure) manifests that the multiple voice system in Squiliq Atayal is a morphosyntactic device encoding not only the alternation of the mapping between f-structure and a-structure but also that of the mapping between i-structure and a-structure. It complements the lexically determined mapping approach in Riesberg’s (2014) to take the mapping with i-structure into consideration. Broadly speaking, grammatical voice is merely one way to encode the alternation of correspondences between representations. Moreover, the linear order of discourse functions displayed by the general mapping unveils an intriguing ideal ranking of i-structural constraints: that is, the Aboutness-Last Constraint ranks over the Given-First Constraint. This ranking more or less restricts the word order in c-structure via mapping and results in the formation of the unusual unmarked word order VOS in linguistic typology.
7. 

Correspondences (II):
Grammaticalized Topic

---

Power exists in one's capacity to link his will with the purpose of others, to lead by reason and a gift of cooperation.

-- Woodrow Wilson

7.1 Overview

This chapter focuses on the marked clauses containing grammaticalized topics (GTOPs)—that is, topic constructions. There are two major parts in this chapter. The first part concerns the empirical facts of different topic constructions, including topicalization, left-dislocation, dangling topic constructions, and utterance-final particles, and their properties are presented with details. These empirical facts further lead us to the second part of this chapter, which concerns the discussion about important theoretical issues, including whether the existence of grammaticalized discourse functions in f-structure is necessary, what c-structural position GTOPs are located in, whether the relation between GTOPs and their comment parts is established on functional control or anaphoric control, what lexical properties the topic marker *ga* possesses, and most important of all, what discourse functions in i-structure GTOPs correspond to.

7.2 Topic Constructions

GTOPs areaccommodated in several topic constructions in Squilq Atayal,
including topicalization, left-dislocation and dangling topics. They are introduced one by one in this section. In addition, it will be shown that the topic marker *ga* has an extended use, being a utterance-final particle.

### 7.2.1 Topicalization

Topicalization is a traditional generative term indicating the movement of a constituent from its original position to the topic position. Here it is used without any connotation about movement but to emphasize that topicalized constituents play a double role in their own clause: they are not only marked topics but also arguments of the predicate at the same time. Rau’s (1992) grammar provided a brief but accurate description of topicalization in Squilq Atayal: “[t]he subject may also be followed by the particle *ga’* and a pause to form a topic apart from the rest of the sentence. A topic introduces a sentence by telling what the sentence is about and always comes first in the sentence. (p.136)” This description pointed out four properties of topicalization. First, there is an omissible topic marker *ga* following a GTOP. Second, there is a pause between the GTOP and the rest of the clause. Third, GTOPs are in the clause-initial position. Last, GTOPs are associated with subjects.

The first and third properties have been noted in passing in Chapter 3, as illustrated in (218) below:

```
(218) pagay qani (ga), kguh-an na’ ngta’ Ø.
    rice  this Top scatter-LV Obl chicken Ø
‘(Speaking of) the rice, (it) was scatterd by the chicken.’ <E>
```

In the above example, the GTOP *pagay qani* ‘the rice’ is placed in the initial position of this sentence. It is followed by the omissible topic marker *ga*, and whether the topic marker is omitted does not affect the grammaticality and meaning of this sentence. In addition, there is a missing subject inside the comment part *kguhan na’ ngta’* ‘was
scattered by the chicken’, as marked by the null symbol Ø. In terms of thematic relation, the missing subject is the patient of the main verb, and it referent is identical with that of the GTOP *pagay qani* ‘the rice’.

As for the second property, the intervention of a pause between a GTOP and its comment part means that they belong to different intonation contours. This is illustrated in Figure 7-1 below.

### Figure 7-1 *Intonation of Topicalization*

The above intonation pattern comes from a clause in a naturalistic narrative. It shows that the GTOP *soliq qasa ga* ‘(as for) that woodgun’ is separated from the rest of the clause (comment) by a pause lasting around 0.5 second (circled in this figure). In addition, the content word in this GTOP (*soliq* ‘woodgun’) has the highest pitch in the whole clause, particularly on its stressed (final) syllable.

The omission of the topic marker *ga* does not affect the existence of a pause, as presented below:
The sentence in Figure 7-2 is a fragment of a long narrative. Although there is no topic maker *ga* following the GTOP *qpring ru*’ *ryuk* ‘(as for) nutgall and walnut’, it is still marked by a 0.25 second pause intervening between topic and comment. It is this pause that differentiates GTOPs from GFOCs. Again, the final syllable of this GTOP bears the highest pitch in the whole clause.

The fourth property means that topicalization in Squilq Atayal follows the subject-only constraint: only subject can be topicalized. If an argument other than subject is topicalized, the sentence will become ungrammatical:

(219) *bzyok qnyun ga, m-etaq lalaw qu’ yukan.

pig mountain Top AV-stab knife Nom Yukan

Intended for ‘(Speaking of) the wild boar, Yukan stabbed (him) with a knife.’ <E>
(220) *lalaw ga, m-etaq bzyok qnhyun qu’ yukan.
    knife Top AV-stab pig mountain Nom Yukan
Intended for ‘(Speaking of) the knife, Yukan stabbed a wild boar (with it).’<E>

The above two sentences contain main verbs carrying the AV prefix m-, but they are both ungrammatical because the GTOP bzyok qnhyun ‘wild boar’ in (219) is the object of the sentence and the GTOP lalaw ‘knife’ in (220) is the instrumental adjunct of the sentence. Neither object nor instrumental adjunct complies with the subject-only constraint in topicalization.

The same constraint applies to the other three voices (PV, LV and CV) in Squilq Atayal as well. In PV sentences, only the grammatical subject, which is also the patient of the action, can be topicalized, as exemplified by the grammaticality of (221) and (222):

(221) qulih ga, niq-un tali’.
    fish Top eat-PV Tali
‘(Speaking of) the fish, (it) will be eaten by Tali.’<E>

(222) *tali’ ga, niq-un qu’ qulih.
    Tali Top eat-PV Nom fish
Intended for ‘(Speaking of) Tali, the fish will be eaten (by him).’<E>

The above sentences have in common their thematic functions and grammatical functions: tali’ is the actor and the oblique, and qulih ‘fish’ is the patient and the subject. Their difference, however, lies in which argument is preposed as the GTOP. The subject is topicalized in (221) while the oblique is topicalized in (222). It follows that even the actor cannot be topicalized when it is not subject.

The subject-only constraint on topicalization is attested once more by the following LV and CV sentences:
(223) slaq qani ga, qniq-an qulih tali’.
   wet.field this Top eat-LV fish Tali
   ‘(Speaking of) this wet field, (it) was the place where Tali ate fish.’ <E>

(224) *tali’ ga, qniq-an qulih qu’ slaq qani.
   Tali Top eat-LV fish Nom wet.field this
   Intended for ‘(Speaking of) Tali, this wet field was the place where (he) ate fish.’ <E>

(225) *qulih ga, qniq-an tali’ qu’ slaq qani.
   fish Top eat-LV Tali Nom wet.field this
   Intended for ‘(Speaking of) the fish, this wet field was the place where Tali ate it.’ <E>

(226) qway ga, s-qaniq qulih tali’.
   chopstick Top CV-eat fish Tali
   ‘(Speaking of) chopsticks, (they) were used by Tali to eat fish.’ <E>

(227) *tali’ ga, s-qaniq qulih qu’ qway.
   Tali Top CV-eat fish Nom chopstick
   Intended for ‘(Speaking of) Tali, the chopsticks were used (by him) to eat fish.’

(228) *qulih ga, s-qaniq tali’ qu’ qway.
   fish Top CV-eat Tali Nom chopstick
   Intended for ‘(Speaking of) the fish, the chopsticks were used by Tali to eat (it).’ <E>

The verbs in (223), (224) and (225) carry the LV suffix -an, and those in (226), (227) and (228) carry the CV prefix s-. Therefore, slaq qani ‘this wet field’ in the first three examples is both the location (thematic function) and the subject (grammatical function); similarly, qway ‘chopstick’ is both the instrument (thematic function) and the subject (grammatical function) in the last three examples. In all of the above six examples, tali’ is the actor and qulih ‘fish’ is the patient; neither of them are the subject. If what is
topicalized is the subject, like *slaq qani* ‘this wet field’ in (223) and *qway* ‘chopstick’ in (226), the sentences will be grammatical; in contrast, if non-subjects (*tali* and *qulih* ‘fish’) are topicalized, like (224), (225), (227) and (228), the sentences will become ungrammatical.

In conclusion, by observing topicalization in different voices in Squliq Atayal, we provide solid evidence to prove that subject is “the only core argument” which can be topicalized in this language.\(^5\)

### 7.2.2 Left-Dislocation

Left-dislocation is a construction rarely researched in previous literature since it is like topicalization in many ways. The only difference between topicalization and left-dislocation lies in the fact that the former has a gap of argument in the comment part whereas the latter does not have any gap but contains a resumptive pronoun coreferential with the GTOP. This is illustrated below:

\[(229)\] yukan= su’ ga, betunux yal syasin= nya’.

Yukan= 2SGen Top beautiful very photo= 3SGen

‘(Speaking of) your Yukan, his photo is very beautiful.’ <N>

In the above example, the resumptive pronoun -*nya*’ in the comment part is co-referential with the GTOP *yukan= su’* ‘your Yukan’, expressing the possessor-possessed relationship between the GTOP and *syasin* ‘photo’. A left-dislocated GTOP is also followed by the omissible topic marker *ga* and placed in the sentence-initial position. What is different from topicalization is that the comment

\(^5\) Nominal predicates are not discussed in this section. Since they contain only one argument, subject, they undoubtedly comply with the subject-only constraint, as shown below:

(i) sayun ga, kneril na’ tali’.

Sayun Top woman Gen Tali

‘(Speaking of) Sayun, (she is) Tali’s wife.’ <E>
part \textit{betunux yal syasin=nya} ‘his photo is very beautiful’ itself is a complete proposition without any missing argument. Despite the controversy over whether topicalization and left-dislocation are different in syntax, semantics and pragmatics nor not and what these differences really are in English (Gregory & Michaelis 2001, Prince 1984), they are quite similar in Squliq Atayal.

In addition to the accompany of omissible topic marker \textit{ga} and the location in the sentence-initial position, one of their similarities is that a left-dislocated GTOP and its comment part belong to different intonation contours, which are separated by a pause, as shown in Figure 7-3 below.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure73.png}
\caption{Intonation of Left-Dislocation}
\end{figure}

In the above figure, the GTOP \textit{kun hozil qani ga} ‘(speaking of) this dog of mine’ is separated from the rest of the clause by a pause lasting 0.5 second.

Although resumption is an important grammatical means to connect a GTOP topic and its comment, it is restricted in the use of realizing the possessor-possessed
relationship in semantics. Put differently, GTOPs cannot be co-referential with resumptive pronouns in the argument position, as illustrated in the following examples from grammaticality judgment tests:

(230) *watan, ga, b<n>hy-an na’ sayun **hiya’i**.
    Watan Top <Perf>hit-LV Obl Sayun 3SNeu
    Intended for ‘(Speaking of) Watan, he was beaten by Sayun.’ <E>

(231) *watan, ga, <n>niq-an=**nya’i** quilih.
    Watan Top <Perf>eat-LV=3SObl fish
    Intended for ‘(Speaking of) Watan, the fish was eaten by him.’ <E>

When the GTOP *watan has the same referent as the pronominal subject *hiya’ ‘he’ in (230), the whole sentence becomes unacceptable. Even when the resumptive pronoun is located in an oblique position, like *-nya’ ‘by him’ in (231), the co-reference between the GTOPs and the pronominal arguments still causes ill-formedness. These two examples show that Squliq Atayal lacks the typical case of left-dislocation in which resumptive pronouns are able to function as core-arguments in the norm of grammar since informants strongly reject these examples in the grammaticality judgment tests. However, counterexamples exist in the naturalistic data. How to interpret this conflict between grammaticality judgment and naturalistic data will be discussed in Chapter 9.

### 7.2.3 Dangling Topic Constructions

In dangling topic constructions, the relationship between a GTOP and its comment part looks looser than that of left-dislocation. It is built on the aboutness relationship or the adverbial modification, not on resumption. However, they are like topicalization and left-dislocation in terms of the topic marker, the intonation pause and the sentence-initial position.
7.2.3.1 Aboutness Topics

Aboutness topics usually refer to Chinese-style topics (Li & Thompson 1981, Tang 1977, Tsao 1979, 1990). Cheng (1989) stated that “aboutness topics are never linked to a gap. Instead, they usually denote a superset of the referent of some other NP in the sentence or a possessor-possessed relationship (p.44).” This section further narrows down this definition: aboutness is the relation between GTOPs and their gapless comments which can only be established semantically or pragmatically without resorting to syntactic means. In other words, resumption observed in left-dislocation is excluded from the discussion on aboutness topics in this section. The following are typical examples from Mandarin Chinese (Chao 1968, Huang, Li & Li 2009, Li & Thompson 1981, among many others):

(232) zhe-ke shu, yezi hen da.
   this-CL tree leaf very big
   ‘(As for) this tree, (its) leaves are big.’

Shi (2000) claims that GTOPs and their comments in Chinese-style topic constructions have a structural relationship, contrary to most of traditional analyses. Some of his arguments and his grammaticality judgments are somewhat controversial. However, I do agree that the relationship between external topics and their comments are not as loose as it was described in some previous studies. Aside from the topic analysis, Xu (2003) made a purely theoretical proposal that Chinese is a language permitting multiple subject because of recursive IPs and that Chinese-style topics are subjects in nature. This proposal is not widely accepted.

Another type of aboutness topics is like the following famous example:

(i) nei-chang huo, xingkui xiaofangdui lai de kuai.
   that-CL fire fortunately fire-brigade come DE fast
   ‘(As for) that fire, fortunately the fire brigade came fast.’

According to Shi (2000), this sentence has two meanings: one is ‘(As for) that fire, fortunately the fire brigade came fast, (otherwise it would have killed many people).’, and the other is ‘(As for) that fire, fortunately the fire brigade came fast, (otherwise we would all have been burnt to death at that time).’ The first meaning is licensed by the adverbial xingkui “fortunately”. To my interpretation, the prima facie comment is not a real comment but the adverbial clause of the unspoken matrix clause (the real comment). In the second meaning, the external topic is a sentential adverbial.
In (232), the external topic *zheke shu* ‘this tree’ is the possessor semantically associated with the possessed *yezi* ‘leaf’ in the rest of the clause. Compared with the English translations, Mandarin Chinese does not need the existence of genitive pronouns to establish this semantic relationship. In (233), the GTOP *shuiguo* ‘fruit’ is the semantic superset of the phrase *xiangjiao* ‘banana’.

Squliq Atayal is like English but unlike Mandarin Chinese since GTOPs cannot express the possessor-possessed relationship without resumptive pronouns:

(234) *watan ga, m<iihiy sayun qu’ yaya’*(=nya’).

Watan Top AV<Perf>hit Sayun Nom mother=3SGen
‘(Speaking of) Watan, his mother hit Sayun.’ <E>

(235) *watan ga, m<iihiy hozil*(=nya’) qu’ sayun.

Watan Top AV<Perf>hit dog=3SGen Nom Sayun
‘(Speaking of) Watan, Sayun hit his dog.’ <E>

The above two sentences show that it is ungrammatical for sentences to lack the genitive pronoun -nya’ ‘his’ to express the possessor-possessed relationship between the external topic *watan* and the possessed, *yaya* ‘mother’ and *hozil* ‘dog’. To wit, the possessor-possessed relationship can only be established by means of resumption.

There are two types of aboutness topics in Squliq Atayal. The first type of aboutness topics has a twin double in comment.

(236) *nanu’, knan uzi ga, baq=saku’ kbalay uzi*.

Par 1SNeu too Top AV.know=1SNom AV.make too
“Well, (speaking of) me as well, I know how to make (it), too.” <N>
(237) aw isu’ kyaqah qani ga, nyux=su’ m-zyu laqi. yes 2SNeu devil this Top Aux.Prog=2SNom AV-trick child “Right, (speaking of) you this devil, you are fooling the child.” <N>

(238) sami’ yaya’ he ru’ yaba he ga, ini’=sami’ <n>bah mha yaqih 1PNeu mother Emp and father Emp Top Neg=1PNom <Perf>change thus evil lungan=myan. heart=1PGen “For us parents, we did not make our hearts in such a pitiless way.” <N>

(239) pila’ qani ga, pila’ mama’, ulungsu’ m-aki’ pila’ money this Top money uncle luckily AV-exist money ka’ sbly-an mama’ qani. Cmp bequeath-PV uncle this ‘(Speaking of) this money, Uncle’s money, luckily there is the money Uncle bequeathed.’ <N>

In the first three examples, all of the topics are first/second person free pronouns, which have the same referents with the first/second personal bound pronouns in the comment part. Similarly, the GTOP in (239) pila’ ‘money’ is repeated in the comment part but it is modified by a relative clause. The relation between GTOPs and their twin doubles in the comment part requires the hearer’s interpretation since the first and second person pronouns and common nouns never serve as anaphors. Note that these doubles are all subjects in the comment part.

The second type of aboutness topics implicates the superset/superordinate-subordinate relationship. That is to say, GTOPs are a semantic superset of the argument in the comment part:

(240) qhoniq ga, sy-on=maku’ qu’ mtalah na’ qhoniq. tree Top like-PV=1SObl Nom red Gen tree ‘(Speaking of) trees, I like red trees.’ <E>
(213) qulih ga, m<n>qniq=saku’ krahu’ na’ qulih.
fish Top AV<Perf>eat=1SNom big Gen fish
‘(Speaking of) fish, I have eaten big fish.’ <E>

(241)
(242) laqi ga, sy-on=maku’ qu’ laqi mlikuy.
child Top like-PV=1SObl Nom child man
‘(Speaking of) children, I like boys.’ <E>

In the above three sentences, the external topics—qhoniq ‘tree’ in (240), qulih ‘fish’ in (241), and laqi ‘child’ in (242)—are the superset of the subjects—mtalah na’ qhoniq ‘red trees’, krahu na’ qulih ‘big fish’ and laqi mlikuy ‘boys’ respectively. These sentences are all acceptable to native speakers, but if the head noun is removed, the whole sentence will sound more natural due to no repetition. Take (240) for example. Without the head noun qhoniq ‘tree’ in the subject, according to my informants, it will be rhetorically better, as the following:

(243) qhoniq ga, sy-on=maku’ qu’ mtalah.
tree Top like-PV=1SObl Nom red
‘(Speaking of) trees, I like red ones.’ <E>

This section shows that external topics in Squiliq Atayal can have the “aboutness” relationship with the subjects inside the comment parts. The aboutness topics are further classified into the twin double type and the superset type. Like English, the possessor-possessed relationship can only be obtained with the existence of resumptive pronouns.

7.2.3.2 Scene-Setting Topics

Scene-setting topics are called stage topics in Erteschik-Shir (2007), providing the information of the particular situation about which they are asserted. The relationship between scene-setting topics and their comments is built on adverbial modification.
Like left-dislocation and aboutness topics, the comment part of a scene-setting topic is a complete proposition without any missing argument. Grammaticalized scene-setting topics are prevailing cross-linguistically, such as Mandarin Chinese (Tsao 1990), Zhuang (a Tai-Kadai language) (Wei, He & Lo 2011), Achang (a Tibeto-Burman language) (Shi 2009), Nigerian Arabic (a Semitic language) (Owens 1993), among others. The following example is from English:

(244) Outside the door, there’s a cat. (Erteschik-Shir 2007:17)

Likewise, GTOPs can function as adverbials in Squliq Atayal, as illustrated in the following examples.98

(245) suxan ga, m-usa sincik temu. (s. m. Hsiao 2004:106)
   tomorrow Top AV-go Hsinchu Temu
   ‘Tomorrow, Temu will go to Hsinchu.’

(246) hira’ sasan ga, m-tyaw (qu’) temu. (s. m. Hsiao 2004:106)
   yesterday morning Top AV-work Nom Temu
   ‘Yesterday morning, Temu worked.’

(247) raral ga, m-aki’ beh yatux, krahul yatux qu’
   past Top AV-exist near high.mountain AV.come high.mountain Nom
   mama’ qasa.
   uncle that
   ‘A long time ago, that uncle lived in high mountains, and came from high
   mountains.’ <N>

98 If the transformation approach is adopted, as Hsiao’s (2004) research on adverbials in Squliq Atayal, these adverbials (especially temporal adverbials) will be analyzed as GTOPs moving out of the comment part and they are internal topics. According to Hsiao (2004), temporal adverbials have seven possible positions in a sentence: topicalization, clefting, TVOS, VTOS, VOTS, VOST (T = temporal adverbial; V = verb ; O = object ; S = subject). In order to explain why only temporal adverbials and subject can be topicalized, she adopts the proposal by Tsai (2004): the head of CP carries an uninterpretable T-feature (tense-feature), and only temporal adverbials and subject carry T-feature. By means of movement, either subject or temporal adverbials can be raised to check these T-features off.
In (245), (246) and (247), temporal adverbials suxan ‘tomorrow’, hira’ sasan ‘yesterday morning’ and raral ‘past’ are all followed by the topic marker ga. In addition, frequency adverbials like kruma’ ‘sometimes’ in (248) and transitional adverbials yasa ‘thus’ in (249) can be GTOPs as well.\textsuperscript{99} What looks more confusing at first sight is the example in (250), in which the comment only has a nominal predicate ngasal balay ‘good house’ without any overt subject. It is highly likely that we interpret the GTOP tanux-nya’ ‘its front yard’ is interpreted as the subject of this nominal predicate until we realize that “its front yard” cannot be “a good house” as shown in (251). As a matter of fact, the subject of the nominal predicate refers to a house in previous discourse, and the GTOP tanux-nya’ ‘its front yard’ is the reason of this sentence. These adverbials are grammaticalized scene-setting topics, which employ the tempo-spatial information in

\textsuperscript{99} There is no doubt that kruma’ ‘sometimes’ have positions other than GTOPs, as shown below:

\begin{itemize}
  \item[(i)] ...ru’ wal si gluw m-huqil kruma’. (Y&Y1991)
\end{itemize}

In the above sentence, it is apparent that frequency adverbials can follow verbs. Because the subject is a zero anaphor (or pro), so it is not clear whether frequency adverbials precede or follow the subject.
the process of information transmission to attract the hearer’s attention and further to proceed with the event-reporting.

Moreover, the topic marker *ga* can connect a clause, not just a phrase, with another clause to constitute a complex sentence; meanwhile, the clause it is attached to functions as an adverbial subordinate clause. In other words, the topic marker *ga* seems like a subordinator, the interpretation of whose meaning relies upon the context:

(252) nanu’ a, sbk-an=nya’ lozi ga, s-on sqani m-aras.
Par Par split-LV=3SObl again Top go-PV this AV-pull
‘Well ah, after it is split again, pull (it) from this.’ <N>

(253) cyux krgyax ga, blaq yal m-emaw, a, lokah yal.
exist mountain Top good very AV-seem Par AV.strong very
‘Because (the cabbages) grow in the mountains, (they) are very good, (they) seem, ah, to be very healthy.’ <N>

(254) m-usa pabu’ ga, m-aki’ qu’ m-htuw zyaw ay. (Rau 1992:190)
AV-go shoot Top AV-exist Nom AV-appear thing Excl
‘If (they) go hunting, problems will arise.’

(255) ana laqi qu’ tali’ ga, ini’ thoyay m-aniq qitun.
even child Nom Tali Top Neg AV.able AV-eat corn
‘Although Tali is young, (he) was not able to eat any corn.’ <N>

In the above complex sentences, those adverbial subordinate clauses are joined with the main clause with the topic marker *ga*. This is just like what has been shown in the above examples of scene-setting topics. To put it differently, this marker renders those phrases and clauses the status of being adverbials. Again, by resorting to the given or tempo-spatial information, the topic marker *ga* marks phrases or clauses as scene-setting topics to unfold the following event (new information). That is why

---

100 Tense is anaphoric; if the preceding discourse fails to provide any reference time, a temporal adverbial
these adverbial clauses are obligatorily located in the sentence-initial position. Unlike those nominal topics discussed in Section 7.2.1 and 7.2.2, adverbial topics do not initial a discourse topic for the following discourse. As Givón (1990:846) says, “…only the strictly-anaphoric property of ‘givenness’ could possibly apply to pre-posed ADV-clauses. But in contrast with prototypical nominal topics, preposed ADV-clauses are seldom recurrent—either cataphorically or anaphorically.”

Furthermore, the topic marker ga can follow adverbials like ini ‘not’, ana ‘even’ and si ‘if’ (L. Huang 1993, Rau 1992, Su 2004), as illustrated below:

(256) sy-an=nha’ biru, ini’ ga sy-an=nha’ kinenhin. put-LV=3PObl book Neg Top put-LV=3PObl souvenir ‘They put books (on it), or they put souvenirs (on it).’ <N>

(257) ana ga zyaw mlikuy kneril ga si ryax m-htuw uzi. even Top thing man woman Top still always AV-happen too ‘But still questions of men and woman come up all the time.’ (Egerod 1999:12)

(258) nanu’ si ga, s’ang-un, ini’ ltux. Par if Top shout.down-PV Neg AV.bark ‘Well, if (the dog) is shouted down, (it) will not bark.’ <N>

To sum up, the topic marker ga can follow phrases or even clauses to form scene-setting topics functioning like adverbials in the sentence, the meaning of which include time, frequency, reason, condition, concession, and so forth.

---

is required (Hinrichs 1986). In other words, as discussed in Horstein (1990), preposed temporal adverbials systematically introduce a new reference time. Based on the interaction between temporal adverbial phrases/clauses and quantificational adverbs, de Swart (1999) proposed that preposed temporal adverbials (phrases or clauses) are topics, “providing the temporal frame for the main clause. (p.359)” The data in Squliq Atayal not only square with this analysis but also provide it with solid evidence since preposed temporal adverbials are all marked with the topic marker ga.

101 This is the same in Mandarin, a topic-prominent language in Li and Thompson (1976). Tsao (1988) analyzes adverbial clauses in Mandarin Chinese as topicalized complex nominal phrases.
7.2.4 Extended Usage

Among all the uses of the topic marker ga, the most deviant one is the status of being an utterance-final particle, which follows an utterance (usually a sentence) to convey various “emotions” or “attitudes”.\(^\text{102}\) Although they have nothing to do with truth-conditional semantics or well-defined grammatical functions, they often play a crucial role in the interpretation of the utterances in which they occur. The utterance-final particle ga is optional in the following sentences:

\begin{align*}
(259) & \text{kyal-aw}=\text{misu’ ga.} \\
& \text{talk-PV.Proj=1SObl.2SNom Par} \\
& \text{‘Let me tell you.’ } <\text{N}> \\
(260) & \text{bsyaq cikay la laqi=nya’ ga?} \\
& \text{long.of.time a.few Par child=3SGen Par} \\
& \text{‘Is his child here for a long time?’ } <\text{N}> \\
\end{align*}

Nevertheless, even native speakers cannot clearly elaborate what the difference is with the attachment of ga here. Detailed analysis is far beyond the scope of the present thesis, yet we can observe some facts it show to us. First, it can be attached to projective, interrogative, and declarative sentences, as shown in (259) and (260). Second, this utterance-final particle carries a rising in the intonation contour. This is illustrated in the following figure:

\(^{102}\) Utterance-final particles can be further analyzed as “discourse markers with only very general context-free properties, which when used in different discourse contexts, may interact with the specific contextual features and help to derive the specific interpretations of the utterances to which they are suffixed (Ing Cherry Li 1999:4).” Like Squliq Atayal, some utterance-final particles in both Old Chinese and Mandarin Chinese can function as topic markers as well (Chu 1993, 1998, 2010, Li & Thompson 1981), and the relation between these two uses is inferable via pragmatic principles (Chu 1998, 2010). The well-known topic marker wa in Modern Japanese also functioned simultaneously as an utterance-final particle in the history of the Japanese language (Shibatani 1990). However, it is not universal that topic markers function as utterance-final particles at the same time. For example, as pointed out in Guo (2008), this association does not stand in Jin, one of the Chinese languages.
Declarative sentences usually have a falling intonation at the end, but the above figure shows that it is not the case for those suffixed with the utterance-final particle \( ga \), the rising pitch of which brings a steep pitch movement vis-à-vis the rather flat intonation contour of \textit{mama-mu'} ‘my uncle’. Acoustic prominence is typically made up of (1) either steeper pitch movement or change of direction; (2) extended duration, and (3) slightly greater intensity (Halliday & Greaves 2008:103). This rising in intonation represents acoustic prominence, definitely attracting the hearer’s attention. It is undoubted that this attraction does not aim to highlight the particle itself since it is a function word, but to highlight the whole sentence so as to evoke the hearer’s participation/response in communication.

103 In this sentence, there is an obvious slip of tongue. The speaker originally wanted to say \textit{k mama-mu’} “my deceased uncle” but she erroneously started the word with \textit{kmu}. That is why this wrong word was stopped soon and a very short pause (0.1 second) was inserted.
7.3 Grammaticalized Discourse Functions

As mentioned in the literature review in Chapter 2, the information-structural notions in LFG developed from Bresnan & Mchombo’s (1987) proposal for TOP as a different grammatical function from SUBJ in f-structure. TOP and FOC were grammaticalized discourse functions serving to express relations related to discourse representation (Bresnan 2001). Their existence however could not explain predicate focus in Russian, so King (1997) argued that i-structure as an independent representation was needed to explain discourse-pragmatic phenomena (particularly on morphosyntax). This arouses a question: do we still need grammaticalized discourse functions in f-structure since we already have i-structure?

The position we take here resembles that of Alsina (2008) and Asudeh (2012): some particular overlay functions have to remain in f-structure but their labels have to be changed. The original labels TOP and FOC are syntactically-wrapped information-structural notions, the existence of which will blur our understanding of the interactions between f-structure and i-structure. As we will present in the later sections, TOP and FOC in f-structure may not be related with the discourse functions named with topic and with focus respectively in i-structure. To be more specific, it will be proved later that in Squliq Atayal, one of the discourse functions which GFOC corresponds to in i-structure is contrastive topic (CSTOP). If GFOC is analyzed as having the grammatical function FOC in f-structure, the label FOC will be misleading. Her (1985-1986, 1991) reminded linguists of the importance of distinguishing the syntactic notion of topic and the discourse notion of topic.104 The overlay grammatical functions should be neutral to information-structural notions. The label used in Alsina (2008) is

---

104 Xu (2005[2003]) sketchily categorized Chinese linguists’ studies on topic & subject into four types: (i) topic in syntax; (ii) subject in the syntax and topic in discourse; (iii) subject plus topic in syntax and topic in discourse; (iv) subject with topic feature in syntax and topic in discourse. This categorization shows the complexity of the relations between f-structure and i-structure.
OP (operator). However, as criticized in Asudeh (2012), it leads to confusion with pre-existing semantic terminology so that UBF (unbounded dependency function) is suggested in Asudeh (2012). In our opinion, UBF is still not good enough since this label is too specific to dependency-related constructions and seemingly exclusive of others. Here we suggest that GOF (grammaticalized overlay function) is a better label since it is not only neutral to information-structural notions but also unspecific to any particular constructions. In the present thesis, the original label TOP is replaced by GOF1 and FOC by GOF2 in f-structure. The numbers are merely for convenient distinction and have nothing with priority or ordering.

All grammaticalized overlay functions must functionally control or anaphorically bind ordinary grammatical functions, like SUBJ, OBJ or ADJ, following the Extended Coherence Condition proposed in Fassi-Fehri (1988), which is based on the Coherence Condition mentioned in Chapter 2. It is paraphrased in (261).105

\[(261) \textbf{Extended Coherence Condition} \text{ (Falk 2001:64)}\]

All functions in an f-structure must be incorporated into the semantics. Argument functions are subject to the Coherence Condition. Overlay functions must be identified with arguments or adjuncts. Adjuncts must be in f-structures containing PREDs.

At least in Squilq Atayal, GOF1 and GOF2 may be subtypes of a general GOF since the targets which they are identified with under the Extended Coherence Condition are in complementary distribution. Specifically speaking, as we will present in later sections, the targets of GOF1 are SUBJ and ADJ while that of GOF2 is PREDLINK. It will be more economical for our theory to lump GOF1 and GOF2 together as one grammatical function. However, for the sake of clarity and convenience, we still make a distinction between

105 C. Huang (1989) claimed that semi-transitive verbs in Mandarin Chinese, such as nashou ‘be good at’, subcategorize TOP. However, Her (1991, 2010) argued against the idea of subcategorized TOP; in his analysis, the verb nashou still subcategorizes object but it requires that the subcategorized an object carry [FRAME +] (i.e. [+ GIV] in our terminology) and be placed out of the predicate carrying new information.
GOF1 and GOF2 in the present thesis.

As for how many GOFs can be licensed simultaneously in a clause, it is a language-particular property, and this property may change through time. For instance, the number of preposed elements and their types has gradually decreased in the history of the English language (Ge 2008, Los 2012).\textsuperscript{106} It is possible that the number and the type of GOFs in a clause are restricted by a language-particular constraint. Anyway, after the presentation of our position, the subsequent sections will continue the discussion of marked clauses in Squliq Atayal. For the convenience of discussion, “grammaticalized topic (GTOP)” and “grammaticalized focus (GFOC)” will continue to be used to refer to the topic/focus expressions in marked clauses but not grammaticalized discourse functions in f-structure. The grammaticalized discourse functions they carry in f-structure are GOF1 and GOF2 respectively.

7.4 Theoretical Issues

GTOPs are discourse topics substantiated in marked constructions. It will be argued in this section that they are CP-adjuncts in c-structure and switch topics or scene-setting topics in i-structure no matter whether they are coreferential with empty subjects or not. At the end, it is shown that the position of GTOPs is determined by the cooperation of f-structure and i-structure.

7.4.1 Syntactic Structure

In the discussion on f-structure in Chapter 3, temporal adverbials serving as GTOPs are tentatively placed in the CP-specifier, just like \textit{wh}-words in English. This CP-specifier analysis however does not fit into Squliq Atayal. Since only CP and IP are

\textsuperscript{106} Diachronically speaking, the sentence-initial position, i.e. the CP or IP-adjuncts, developed gradually from a multifunctional position to a more functionally-restrictive position. Los (2012) argues that it brought about the typological swift from bounded to unbounded systems so that the temporal information of every sentence in Present-Day English is not anchored with previous discourse, unlike Modern Dutch and German.
recognized in LFG, the possible loci of GTOPs are CP-specifier, CP-adjunct and IP-adjunct. It is argued in the present thesis that GTOPs in this language are adjuncts to CPs. There are two reasons. First, as mentioned in the first half of this chapter, there is an intonation pause intervening between a GTOP and the rest of the clause, which is different from interrogative clauses in English, commonly recognized as located in the CP-specifier position, where there is no pause between a \textit{wh}-word and the rest of the clause. In addition, the existence of intonation pauses also shows that GTOPs are located outside the intonation contour of their clauses. Thus, the possibility of being IP-adjuncts is excluded.

Second, the existence of multiple GTOPs proves that they are better analyzed as CP-adjuncts than as CP-specifiers. The examples in (262) and (263) both contain two GTOPs preceding the topic marker \textit{ga}: one is an adverbial topic (underlined) while the other is a GTOP coreferential with the grammatical subject in the rest of the clause, i.e. GTOP-SUBJ (in boldface).

(262) \textit{bali nanu’ ga, bnkis rral hiya’,} ga, \textit{m-sli m-sli} \text{Ø}.
\begin{tabular}{p{5cm}p{5cm}}
\text{certainly} & \text{Par Top elder past Emp Top AV-assemble AV-assemble} \text{Ø} \\
\text{‘The reason is that, (speaking of) the forefathers themselves, (they) assembled...’} & \text{<N>}
\end{tabular}

(263) \textit{nanu’ a, pling hiya’,} ga, \textit{bhno-bhnok=nya’} ga, lalu kryu-an \text{Ø}.
\begin{tabular}{p{5cm}p{5cm}}
\text{Par Par nutgall.tree Emp Top Red-softness=3SGen Top easy make-LV} \text{Ø} \\
\text{‘Well, (speaking of) nutgall wood itself, (because of) its softness, (it) is easy to be manufactured.’} & \text{<N>}
\end{tabular}

In (262), the scene-setting topic \textit{bali nanu} ‘certainly’ precedes the GTOP \textit{bnkis rral hiya} ‘the forefather themselves’, which has the same referent with the empty subject marked with \text{Ø} in the comment part (i.e. GTOP-SUBJ). The example in (263), however, has reversed ordering between the scene-setting topic and the GTOP-SUBJ: \textit{pling hiya’}
‘nutgall wood itself’ as the GTOP coreferential with the empty subject precedes the scene-setting topic *bhno-bhno-nya* ‘its softness’. It is shown in the above cases that the ordering between GTOPs is flexible. The ordering flexibility has nothing to do with whether or not the GTOP is coreferential with grammatical subject. That is to say, although we made a tentative distinction between topicalization and other constructions of GTOPs earlier in this chapter, GTOPs of different types are adjuncts in nature. As we all know, adjuncts usually have more flexible word order than arguments in configurational languages. Even if we accept that CPs can be recursive, repeatedly selecting another CP as the complement, we will be faced with the theoretical problem about the nature of the head C: if it is the topic marker *ga*, its omission will force us to recognize the existence of a null topic marker, which violates the economy principle in LFG; if it is not, then what the null head C is will be another problem. It follows that the whole GTOP plus the topic marker is a CP-adjunct. As for what the topic marker *ga* is, it is left to the discussion in Section 7.4.3.2.

Although all GTOPs are the same in terms of their c-structural position, their f-structures vary from type to type. The GTOPs coreferential with null subjects (GTOP-SUBJs) and the dangling GTOPs (GTOP-DG) have different f-structures, as illustrated in the following f-structures:

(264) GTOP-SUBJ (functional control)

```
[GOF1  [⋯]]
[PRED  '⋯']
[SUBJ  [ ]]
```

(265) GTOP-DG (functional control)

```
[GOF1  [⋯]]
[PRED  '⋯']
[ADJ   [ ]]
[SUBJ  [⋯]]
```

GTOP-SUBJs are restricted by the subject-only constraint, simultaneously conveying the two grammatical functions GOF1 and SUBJ, and since there is only one subject in a
clause, there is also only one GTOP-SUBJ in a clause. In contrast, GTOP-DGs are not restricted by the subject-only constraint and behave like appendices, so they convey the two grammatical functions GOF1 and ADJ. Because there can be more than one adjunct in a clause, there can be more than one GTOP-DG in a clause. In spite of these differences, GTOP-SUBJs and GTOP-DGs have one thing in common. That is, they play a dual role of being both a GTOP (GOF1) and a subject/adverbial (SUBJ/ADJ) at the same time. The two grammatical functions GOF1 and SUBJ/ADJ are always identical in their f-structural values. That is to say, for any single GOF1, it must functionally control either SUBJ or ADJ in f-structure. It is formalized as the following equations annotated with the phrase structure rule:

\[(266) \quad CP \to \begin{array}{c} \text{DP} \\ (\uparrow \text{GOF1}) = \downarrow \\ (\uparrow \text{GOF1}) = (\uparrow \{\text{SUBJ} | \text{ADJ}\}) \end{array} \quad CP \uparrow = \downarrow\]

In (266), the phrase structure rule and the equation \((\uparrow \text{GOF1}) = \downarrow\) show that a GTOP is an adjunct to CP in c-structure and carries GOF1 in f-structure. The second equation \((\uparrow \text{GOF1}) = (\uparrow \{\text{SUBJ} | \text{ADJ}\})\) formalizes the subject-only constraint, requiring that GOF1 should equal SUBJ, in the complementary distribution with the circumstance that GOF1 equals ADJ. Despite the interplay between the recursion of CP-adjunction and the subject-only constraint, there are no multiple subjects because the selection of subject must comply with the Coherence Condition that it has to be selected by its predicate (Cf. Chapter 2).

From the theoretical perspective, there is an alternative analysis of the relation between GOF1 and SUBJ/ADJ: that is, it is established on anaphoric control. Anaphoric control has the following three characteristics: first, GOF1 and SUBJ/ADJ are distinct entities with their own values in f-structure; second, SUBJ/ADJ is a pronoun with null phonetic form; third, the co-reference between GOF1 and SUBJ/ADJ comes from the same
index of referentiality. They are sketched below

(267) GTOP-SUBJ (anaphoric control)

\[
\begin{array}{c}
\text{GOF1} \\
\text{INDEX} \quad i \quad i' \\
\text{PRED} \quad \text{‘…’} \\
\text{SUBJ} \quad \text{null} \\
\end{array}
\]

(268) GTOP-DG (anaphoric control)

\[
\begin{array}{c}
\text{GOF1} \\
\text{INDEX} \quad i \quad i' \\
\text{PRED} \quad \text{‘…’} \\
\text{PRED} \quad \text{‘pro’} \\
\text{SUBJ} \quad \text{null} \\
\end{array}
\]

The analysis of anaphoric control is not valid for GTOP-SUBJs and GTOP-DGs in Squliq Atayal, which is supported by the evidence coming from the sentential subject island. This island blocks the co-reference between GOF1 and SUBJ/ADJ, as exemplified below:107

(269) *sayun\textsubscript{i} ga, baq-un=maku’ qu’ [wal m-aniq qulih qani \(\emptyset\)].
Sayun Top know-PV=1SObl Nom Aux.Pt AV-eat fish this \(\emptyset\)
Intended for ‘(Speaking of) Sayun, I know that (she) ate this fish.’
Intended for ‘(Speaking of) Sayun, that (she) ate this fish is known by me.’ (lit.)
\textless E\textgreater

(270) *suxan\textsubscript{i} ga, baq-un=maku’ qu’ [p-aniq qulih qani watan \(\emptyset\)].
tomorrow Top know-PV=1SObl Nom AV.Irr-eat fish this Watan \(\emptyset\)
Intended for ‘Tomorrow, I know that Watan will eat this fish.’
Intended for ‘Tomorrow, that Watan will eat this fish is known by me.’ (lit.) \textless E\textgreater

The GTOP-SUBJ sayun in (269) is co-referential with the null subject \(\emptyset\) across the boundary of sentential subject, which leads to ill-formedness. In the same vein, the GTOP-DG suxan ‘tomorrow’ in (270) is co-referential with the null adjunct \(\emptyset\) across the

---

107 Bresnan (1982), one of earliest research analogizing relativization to topicalization, analyzed relative pronouns as topics under the LFG framework. In Squliq Atayal, relativization obeys the subject-only constraint, just like topicalization discussed in the present thesis (Liu 2004b, 2005), but topicalization is sensitive to the sentential subject island while relativization is not. This difference indicates that topicalization differs from relativization in Squliq Atayal.
boundary of sentential subject, and it causes ill-formedness as well. If the null argument \( \emptyset \) were a pronoun without phonetic form, it should resemble an ordinary pronoun in the ability of referring to the GTOP outside the clause where it exists. The island-sensitivity proves that GOF1 functionally controls SUBJ/ADJ in f-structure.

Even when the null subject is replaced by a resumptive pronoun, as exemplified in (271), the co-reference across the boundary of sentential subject still leads to ungrammaticality.

(271) *sayun, ga, baq-un=maku’ qu’ [wal m-aniq quilih qani hiya’i].
Sayun Top know-PV=1SObl Nom Aux.Pt AV-eat fish this 3SNeu
Intended for ‘(Speaking of) Sayun, I know that she ate this fish.’
Intended for ‘(Speaking of) Sayun, that she ate this fish is known by me.’ (lit.)
<E>

The existence of the resumptive pronoun *hiya’ here is unable to compensate for the violation of the sentential subject island. It indicates that the resumptive pronoun is syntactically inactive, with the co-reference relation established on functional control (Asudeh 2012). That is to say, the resumptive pronoun does not have its own values in f-structure.\(^{108}\)

In contrast, when the resumptive pronoun is genitive, showing the possessor-possessed relationship between GTOP and the argument embedded inside a sentential subject, ill-formedness does not occur. This is exemplified in (272), where the GTOP *sayun is co-referential with genitive pronoun *nya’ ‘her’ across the boundary of sentential subject.

---

\(^{108}\) Asudeh (2012) distinguishes syntactically active resumptives from syntactically inactive ones, the former of which and their binders are distinct entities in f-structure whereas the latter of which and their binders are entities sharing a single value in f-structure.
This is the case of left-dislocated GTOPs (GTOP-LDs). The insensitivity to the sentential subject island proves that the relation between GTOP-LDs and their comment parts is established on binding, not on functional control.

### 7.4.2 Mapping to DFs

After analyzing the c-structure and f-structure of GTOPs, this thesis will investigate in this section what discourse functions in i-structure they correspond to. The GTOPs coreferential with subjects (GTOP-SUBJs) and the left-dislocated/dangling GTOPs (GTOP-LD/DGs) will be analyzed separately.

#### 7.4.2.1 GTOP-SUBJ

The quantitative analysis of topicality in Chapter 5 shows that GTOPs tend to convey important information which can be either new or given in discourse. This tendency do not suffice to identify what discourse functions GTOPs correspond to half because only the cases of transitive clauses are taken into consideration in Chapter 5 in order to study the information arrangement of transitive clauses and half because both givenness and aboutness should be investigated at the same time. Because valence and voice alternation have no bearing on GTOPs, they are lumped together in the following statistical data. In addition, if necessary, we will look into the examples themselves to investigate what discourse function/features they correspond to in i-structure.

Table 7-1 shows the statistical data about givenness and aboutness of GTOP-SUBJs.
There are only two groups: one conveys given and important information (73.4%) while the other conveys new and important information (26.6%). The ratio of them is around 2.8 to 1. The distributive pattern is unlike the scalar hierarchy observed in grammatical subject, and it is also unlike the distribution of grammatical objects, in which all categories have some tokens (Cf. Chapter 6). The analysis of topicality is unable to present that GTOPs all carry the discourse feature [+PROM], which is manifested in their marked word order and independent intonation contour.

Here we would like to argue that all of the information conveyed by GTOP-SUBJs carries [+GIV]; put differently, it is given information with its referent shared by interlocutors, no matter whether it has occurred in previous discourse or not. It is not problematic to equate the first category [+GIV, +ABT] with the discourse function SWTOP in i-structure, which carries [+GIV, +ABT, +PROM]. What may arouse suspicion is the category with new information, i.e. [−GIV, +ABT]. The information that this category contains shows up in the discourse for the first time. Is it really shared between interlocutors? We have to take a close look at the examples, which are illustrated in (273) and (274) below:
Atayal Emp Top AV know AV climb tree
‘(Speaking of) the Atayal people, (they) know how to climb trees.’ <N>

The example in (273) is excerpted from a story about a battle between the Atayal people and the Japanese people. Before this sentence, the informant talked about how bravely his ancestors fought against the Japanese people. Suddenly, he changed the information flow, taking a more objective perspective to comment on this battle, so he made use of the GTOP-SUBJ tayal hiya’ ‘the Atayal people’ in the beginning of (273). Although it is not mentioned in previous discourse, the GTOP-SUBJ itself is semantically generic, referring to all Atayal people as a whole; most important of all, its referent is a piece of given information existing in the world knowledge of both the addressee and the addressee. As for the example in (274), it is excerpted from a narrative about the family history of one of my informants. Before this sentence, the informant stated that no one took care of the family’s old house since the death of her parents. Then, she made use of the GTOP-SUBJ sswe’-mu’ ‘my younger brother’ to change the direction of information flow. Although the referent of this GTOP-SUBJ is not mentioned in previous discourse, it is inferable since the informant has mentioned that her father left eight children to her mother after his death. Therefore, even though 26.6 percent of GTOP-SUBJs seem to convey new information in statistics, they still convey given information to the addressee and the addresser in discourse context.\footnote{This dovetails Rau’s (2000) statistical study that preverbal subjects convey discontinuous information. To be more specific, they convey discourse-new but hearer-old information if the terminology in Birner & Ward (1998) is adopted.}

When one clause with a GTOP-SUBJ is immediately followed by another clause
with a GTOP-SUBJ, these two GTOP-SUBJs display a certain extent of contrast because of the abrupt shift of information flow, as illustrated in (275), an excerpt from a well-known legend about the origin of the Atayal people. In the middle of this legend, the male ancestor and the female ancestor of the Atayal people were trying different ways to have offspring. As shown in example (275), GTOP-SUBJs are both body parts: one is *bling na’ngqwaq* ‘oral cavity’ and the other is *bling na’papak* ‘ear holes’, with boldface added in this example.

(275) **bling na’ ngqwaq** ga kr’us-un=nha’, **bling na’ papak** ga 
hole Gen mouth Top disgust-PV=3PObl hole Gen ear Top 
tpru’ lasa lma. 
AV.stop thus Par
‘As for the oral cavity, (it) was disgusting to them; as for ear holes, (they) were frustrating.’ (Y&Y)

The above two GTOP-SUBJs not only convey some extent of contrast but also serve as listing various items. In fact, this type of GTOPs have been analyzed as contrastive topics in some linguistic studies, such as Vermeulen (2009) on Japanese and Korean. L. Huang & Hayung (2016) also mentions that known information is topicalized for the sake of contrast and comparison. This analysis is not adopted here because contrast in the definition of the present thesis is about whether the information is highlighted against other set members of “the same proposition” (Cf. Chapter 4). In (275), the two sentences have not only different GTOP-SUBJs but also different propositions: one is “for x, x was disgusting to them” while the other is “for x, x was frustrating”. These GTOP-SUBJs do not match our definition on CSTOP. That is to say, the above GTOP-SUBJs are still regarded as conveying the discourse function SWTOP, rather than CSTOP, in i-structure.

Givón (1990) makes a distinction between two types of contrast, referential contrast and thematic contrast when discussing the usage of the conjunct *but* in English.
The former comes only from the reference of noun phrases, as exemplified in (276), while the latter comes from non-referential sources, to my interpretation, predication, as exemplified in (277).

(276) Joe came, but Sally didn’t. (Givón 1990:851)

(277) (Givón 1990:851)
   a. Mary sat down, but she didn’t eat. (switch verb)
   b. Mary sat down, but she Bruce stayed on his feet. (switch subject & verb)
   c. Mary loved lettuce, but loathed spinach. (switch object & verb)
   d. Mary played chess, but Bruch detested board-games.
      (switch subject, object & verb)

(278) Once upon a time there was a king with two sons. The older son expected to take over the kingship…As for the younger, he concentrated on studying philosophy at the University. (s. m. Andrews 2007:149)

Even in a short paragraph with only one switch topic, the switch topic may still bring some extent of contrast, as illustrated in (278), where the switch topic the younger contrasts with the older son against the global discourse background. Lumping all of these examples together under the category of contrast will make it a notion of extreme diversity and obscure our later discussion on the comparison of GTOP and GFOC.

7.4.2.2 GTOP-LD/DG

As mentioned in the last section, both GTOP-SUBJs and GTOP-LD/DGs carry the discourse feature [+PROM] because of their prominence in prosody and word order. In order to identify what discourse functions GTOP-LD/DGs convey, it is necessary to look at the statistical data about the givenness and aboutness of information, as shown in Table 7-2.
Table 7-2 Discourse Features of GTOP-LD/DGs

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>44 (55.0)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>10 (12.5)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>11 (13.7)</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>15 (18.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>80</td>
</tr>
</tbody>
</table>

Although the majority of GTOP-LD/DGs (55.0%) still convey given and important information, unlike what GTOP-SUBJs show, all of the four categories have some tokens, with none empty. This distribution does not mean GTOP-LD/DGs convey four different discourse functions in i-structure. Instead, it is argued that they convey only two discourse functions, that is, SWTOP and SSTOP.

Like GTOP-SUBJs, GTOP-LD/DGs convey given information, which is shared in the knowledge of both the addresser and the addressee. All of the cases with [−GIV] in the above statistical data still carry given information even though the information is introduced to the discourse-pragmatic context for the first time, as exemplified in (279) (repeated from (263)).

(279) nanu’a, pling hiya’ ga, bhno-bhnok=nya’ ga, lalu kryu-an.  
Par Par nutgall.tree Emp Top Red-softness=3SGen Top easy make-LV  
‘Well, (speaking of) nutgall wood itself, (because of) its softness, (it) is easy to be manufactured.’ <N>

In this example, the GTOP-DG bhno-bhnok-nya’ ‘its softness’ introduces the abstract property of nutgall wood into the discourse for the first time, as the reasoning adverbial. Although this information is discourse-new, softness as an abstract property is known to interlocutors. In other words, the categories [−GIV, +ABT] and [−GIV, −ABT] do not exist.
in reality when qualitative evidence is taken into consideration. GTOP-LD/DGs all carry [+GIV] in i-structure. Therefore, GTOP-LD/DGs only have two categories [+GIV, +ABT] and [+GIV, –ABT], which correspond to SWTOP and SSTOP respectively.

The greatest difference between GTOP-SUBJs and GTOP-LD/DGs is the existence of unimportant information, which occupies around 32% of the whole in GTOP-LD/DGs but does not exist in GTOP-SUBJs. That is why GTOP-LD/DGs convey SSTOP. It comes from two situations. The first one is that the referent introduced by GTOPs only serves to introduce another referent sometimes through the possessor-possessed relationship or the superset-subset relationship, eventually replaced by this particular new referent in the information flow and not mentioned anymore in subsequent discourse. This is illustrated in the naturalistic example (280):

(280) **ciriq topan qani** l-ga, cingay balay qu’ bnkis=myan.
    battle Topan this Par-Top AV.many very Nom elder=1PGen
    ‘(Speaking of) the Topan battle, our forefathers were many.’ <N>

The GTOP **ciriq topan qani** ‘the Topan battle’ reintroduces the general discourse topic into the subsequent discourse. Its referent conveys given information but it is not mentioned in the following ten clauses since the subsequent information flow starts to emphasize the referent of the grammatical subject **bnkis-myan** ‘our forefathers’. The second situation comes from the GTOPs representing tempo-spatial information. Tempo-spatial information is always new and not repeated in the subsequent discourse, as exemplified in (281), in which the grammaticalized adverbial topic **misuw qani** ‘right now’ expresses the temporal background, laying the groundwork for the following statement.
Once again, when two clauses with GTOP-LD/DGs are adjacent in discourse, these two GTOP-LD/DGs seem to bring about some extent of contrast, as exemplified in (282).

This example is from a story of family immigration. The GTOPs-LDs yutas-mu’ ‘my grandfather’ and qsuyan-nya’ ‘his elder brother’ are given information, already mentioned in previous discourse. Compared with the similar case of GTOP-SUBJs, the extent of contrast is much weakened since the referents of GTOPs-LDs are only repeated as the possessors, further introducing the grammatical subject lalu-nya’ ‘his name’ in their respective local clause. They are not analyzed as conveying the discourse function CSTOP since their propositions differ in more than one place: “for x, his name is y”. There are two variables in the proposition, which violates the definition of contrastive topics. The feeling of contrast is merely the by-product resulting from listing two SWTOPs within a short discourse span.

Generally speaking, the information conveyed by GTOP-LD/DGs is similar to the information conveyed by GTOP-SUBJs. They all convey given information shared by interlocutors even though some of them are discourse-new. They differ in that there are
some GTOP-LD/DGs which are not repeated anymore in subsequent discourse. Put differently, while GTOP-SUBJs carry [+GIV, +ABT], GTOP-LD/DGs carry either [+GIV, +ABT] or [+GIV, −ABT]. Being characterized by their topic marker and clause-initial position, both of GTOP-SUBJs and GTOP-LD/DGs carry the discourse feature [+PROM] in i-structure. Thus, GTOP-SUBJs correspond to SWTOP in i-structure, and GTOP-LD/DGs correspond to either SWTOP or SSTOP in i-structure. The percentages of SWTOP and SSTOP in our naturalistic data are shown in Table 7-3 below: SWTOP occupies 93% of the whole tokens while SSTOP only occupies 7%.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>DF</th>
<th>Token</th>
<th>Token (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTOP-SUBJ</td>
<td>SWTOP</td>
<td>78</td>
<td>147 (93)</td>
</tr>
<tr>
<td>GTOP-LD/DG</td>
<td>SWTOP</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSTOP</td>
<td>11</td>
<td>11 (7)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>158 (100)</td>
</tr>
</tbody>
</table>

7.4.3 General Mapping of GTOP

Based on the previous discussion of the syntactic structure and discourse functions of GTOPs. GTOPs, their annotated phrase structure rule in c-structure is as illustrated in (283).
The CP-adjoined DP corresponds to a switch topic or a scene-setting topic in i-structure, as annotated in \{(\uparrow_i \text{SWTOP}) = \downarrow, \ (\uparrow_i \text{SSTOP}) = \downarrow\}. The symbols ☃ and ● indicate that their proportions are the range of more than 90% and the range of less than 10%, respectively (Cf. Chapter 5). These topics both require that the information must be given, prominent, as annotated in (\downarrow_i \text{GIV}) = c +, (\downarrow_i \text{PROM}) = c + and (\downarrow_i \text{CONT}) = c − respectively, but it is unspecified whether the information is important in the subsequent discourse. If the CP-adjoined DP corresponds to a SWTOP in i-structure, it must have a positive value of the discourse feature ABT, it must, as shown in the annotation (\uparrow_i \text{SWTOP}) = \downarrow \rightarrow (\downarrow_i \text{ABT}) = c +. If it corresponds to a SSTOP in i-structure, it must have a negative value of the discourse feature ABT, as marked in the annotation (\uparrow_i \text{SSTOP}) = \downarrow \rightarrow (\downarrow_i \text{ABT}) = c −.

The general mapping of GTOPs in (284) is built on the phrase structure rule in (283).
The mapping between c-structure and f-structure is indicated by green lines, the mapping between f-structure and i-structure purple lines, and the mapping between c-structure and i-structure pink lines. The independent type of one-to-many mapping is realized as one mapping line as well as bracketed discourse functions.

7.4.3.1 Prominence

The mapping in (284) illustrates the fact that the CP-adjuncts as GTOPs in Squiliq Atayal have a stronger association with both f-structure and i-structure. On the LFG-OT assumption that c-structure is the output determined by the inputs from f-structure and i-structure, the effect of f-structure is slightly stronger than that of i-structure since GTOPs are specialized for the particular grammatical function GOFI but for two particular discourse functions SWTOP and SSTOP. The discourse functions SWTOP and
SSTOP however have two features in common: [+GIV] and [+PROM]. Thus, the sentence-initial position can be viewed as the position of prominent given information, satisfying both the Given-First Constraint and the Prominence-First Constraint in i-structure. In terms of the construction of GTOPs, Squliq Atayal is a quasi-f-structural language with great influence from i-structure.

In addition, the majority of GTOPs convey SWTOP in i-structure no matter whether they are GTOP-SUBJs or GTOP-LD/DGs. In terms of discourse features, SWTOP is similar with CCTOP since they both are important given information. Their different only lies in whether the information is prominent or not. This similarity opens up the possibility of word order change in diachronic linguistics. Once the frequency of GTOPs is raised, which makes preverbal nominals not unusual any more, the prominence of information will fade out. It has happened to many languages in the world that GTOPs are further grammaticalized into grammatical subjects (Givón 1979a). Squliq Atayal may walk on the same way in the future. Of course, this is just a hypothetical prediction since so far we still can perceive that the discourse features [+GIV] and [+PROM] still play an important role in the determination of CP-adjoined DPs. I-structure has not completely lost its ground yet.

The issue discussed here is closely related with subjectivity and objectivity in cognitive linguistics. People have two ways to observe an entity: one with greater objectivity is purely placing the entity on the stage of attention without any involvement of the observer while the other with greater subjectivity involves both the entity on the stage and the observer (Langacker 1985, 1987). The information introduced by GTOPs involves subjectification since it is not purely presented as the entity on the stage but

---

110 As suggested in Gregory & Michaelis (2001) and Reinhart (1981), topicalization and left-dislocation share the basic function: they are a signal for the listeners to access their mental file on the individual in question, which is promoted to a prominent position in discourse and then is added more information in the following clause. The analysis of GTOPs in this section supports the two previous studies.

111 By studying the word order change of Malay, Cumming (1991) claimed that word order change was not a simple syntactic change affected by the type of topicality but only a part of functional change which involved the change in the distribution of functions over forms.
presented under the addressee’s presumption about the shared knowledge of interlocutors. Thus, from the perspective of cognitive linguistics, we may state that GTOP-SUBJs and sentence-final grammatical subjects differ in that the former involves the egocentric viewing arrangement which locates the addresser together with the entity he/she describes on the stage whereas the latter involves the optimal viewing arrangement which only locates the entity on the stage without any subjective viewpoint of the addresser. In addition, subjective entities are not just true or false in semantics but related with personal attitude (Traugott & Dasher 2002), which is manifested in the discourse features [+PROM] here.

7.4.3.2 Lexical Properties of GA

The last thing needing our clarification is the lexical property of the topic marker ga in Squilq Atayal since the phrase structure rule in (283) does not stipulate its existence. The present thesis suggests that particle is an independent lexical category in this language and that the topic marker ga is a particle. As shown in many of the previous examples, it can be combined with other particles, such as ma and la, forming a bigger chunk which we may view as a particle phrase. A particle phrase is an adjunct to DP, NP, IP or CP in c-structure. What is more important, most of the discourse particles do not have any grammatical function so their existence does not affect the formation of f-structure; on the other hand, they have specified discourse features in their lexical properties. Take the particle ga for example. As shown in (285), its discourse features are [+GIV] and [+PROM].

(285) ga: Party
    (↓, GIV) = +
    (↓, PROM) = +

Through the adjunction to DP/CP, these discourse features are incorporated into the
information conveyed by the nominal/proposition. That is why the phrases which it is adjoined to always convey given and prominent information, such as GTOPs, adverbial clauses and yes-no questions.\footnote{Although yes-no questions aim to inquire the truth value of a proposition (not necessarily so in discourse-pragmatics), the proposition inquired by yes-no questions itself is known to both the addressee and the addressee; otherwise, no addresser would ask a proposition which cannot be understood by the addressee unless there is some pragmatic implicature behind it. That is why the yes-no question can be treated as conveying given information.} The particle \textit{ga} cannot be adjoined to the c-structural position incompatible with [+PROM], such as the object of an AV clause, which is specified with [−PROM]. All of the different uses that the particle \textit{ga} have come from what element it is adjoined to. We may as well state that \textit{ga} serves as a topic maker not because it is specialized as a topic marker in this language but because it is adjoined to DPs. The omission of \textit{ga} in GTOPs is due to the fixed correspondences between c-structure and i-structure, as shown in (283), which enforces the association between syntax and information and makes optional the informational marking of \textit{ga}.

7.5 Summary

This chapter has presented the properties of several topic constructions: topicalization, left-dislocation, and dangling topic constructions. They all share the following three properties of GTOPs. First, GTOPs can be followed by the omissible topic marker \textit{ga}. Second, GTOPs belong to an intonation contour independent from the rest the clause, with the intervention of an intonation pause. Third, GTOPs are located in the sentence-initial position. The differences between these topics constructions mainly lie in how the relationship between GTOPs and their comment parts is built.

In addition, it is suggested in this chapter that the traditional labels of grammaticalized discourse functions in f-structure, TOP and FOC, should be changed to GOF1 and GOF2 respectively so as to remove potential misunderstandings since they may not correspond to topics and foci in i-structure. The neutral terms of grammatical functions in f-structure are beneficial to concentrate linguists’ attention on what
discourse functions in i-structure they correspond to.

A GTOP in Squliq Atayal is argued to be a CP-adjunct in c-structure, and the relation between its grammatical function (i.e. GOF1) and either SUBJ or ADJ in f-structure is mainly established on functional control. Moreover, a GTOP corresponds to SWTOP and SSTOP in i-structure, which indicates that the sentence-initial position is a position of non-contrastive prominent information. At the end of this chapter, the topic marker ga is analyzed as a particle with the two lexical features [+GIV] and [+PROM].
8.

Correspondences (III):
Grammaticalized Focus

I find that a great part of the information I have was acquired by looking up something and finding something else on the way.

-- Franklin P. Adams

8.1 Overview

Following the discussion in Chapter 7, the present thesis continues investigating the correspondences in marked clauses, particularly of grammaticalized focus (GFOC). GFOC is accommodated in pseudo-cleft constructions, the properties of which are introduced in the beginning. Based on these properties and evidence from both quantitative and qualitative methods, several theoretical issues are investigated. First, which approach offers a better analysis of pseudo-clefting in this language, the single-tier one or the double-tier one? Second, what discourse functions in i-structure do GFOCs correspond to? It is proved that not all of them belong to the traditional category of focus (new information). Third, what makes interrogative nominals incompatible with subject? It has something with the discourse features specified in their lexical entries. Last, the issue on how pseudo-clefted GFOCs developed to transitional adverbials is offered a theoretical hypothesis.

The end of this chapter will deepen our understanding of i-structure through the discussion of two previously unmentioned discourse functions, the distinction between primary continuing topic and second continuing topic, and the mapping theory between

243
discourse functions and grammatical functions, the last of which further complements the present LFG framework in which the mapping between f-structure and i-structure is indirect.

**8.2 Focus Constructions**

This section is about focus constructions in Squliq Atayal, basically pseudo-clefting and its related usage of interrogatives and adverbials.

**8.2.1 Pseudo-Clefting**

GFOCs only exist in pseudo-cleft constructions in Squliq Atayal, where an open clause functions as subject and this clausal subject is predicated of a nominal which is co-referential with the missing argument inside it. In English, the open clause is constituted by a headless relative clause, which has no head noun but has a relative pronoun. However, since Squliq Atayal does not have relative pronouns and the open clause looks just like a predicate (Liu 2004b, 2005), the pseudo-cleft construction resembles the word order inversion of subject and its predicate, as described in Rau (1992): “[The verb phrase as] the subject may be introduced by *qu’*, meaning ['']it is the thing or person that…[''] (p.137).” This can be illustrated below:

(286) sy-un=ta’ qu’ cinogan.
    put-PV=1PObl Nom arrow
    ‘Arrows were loaded by us.’ <E>

(287) cinogan qu’ sy-un=ta’.
    arrow Nom put-PV=1PObl
    ‘What we loaded were arrows.’ <E>
(288) nanu’ a, qani qu’ s-betaq.
Par Par this Nom CV-stab
‘Well, what is used to stab is this.’ <N>

(289) ngarux qu’ nyux=nha’ bay kng-un uzi. (Rau 1992:138)
bear Nom Aux.Prog=3PObl very afraid-PV too
‘What make them afraid as well are bears.’

The example in (286) is a sentence with the canonical word order (predicate plus subject), whereas the pseudo-cleft construction in (287) has a sentence-initial nominal predicate cinogan ‘(be) arrows’ and a sentence-final clausal subject syun-ta’ ‘(what) we loaded’, which follows the nominative case marker qu’. If we take a close look at this clausal subject, we will notice that this clause is an open clause which lacks its subject and that the nominal predicate is exactly what the missing subject refers to. Squliq Atayal is a language without copular verbs, so the prima facie analysis is that pseudo-cleft constructions seem to inverse subject and predicate. The other two examples in (288) and (289) present the same fact: what follows the nominative case marker is a clausal subject (sbetaq ‘(what) is used to stab’ and nyux=nha’ bay kngun ‘(what) make them afraid as well’) and what precedes it is a nominal predicate (qani ‘(be) this’ and ngarux ‘(be) bears’). Again, the nominal predicates are co-referential with the missing arguments in these clausal subjects.

Prosodically, unlike topicalization and left-dislocation, there is no intonation pause between the seemingly-preposed subject and the sentence-final open clause. When there is none of case markers and topic markers in a sentence with a preposed nominal, the only way to tell whether this sentence is topicalization or pseudo-cleft is the existence of the intonation pause.113 Look at the following Figure 8-1, and it is shown that there

113 From a diachronic viewpoint, the omission of the nominative case marker qu’ in pseudo-cleft sentences may indicate they are undergoing reanalysis, which results or will result in two constructional actualizations: one is a biclausal pseudo-cleft construction with the nominative case marker, and the other is a monoclausal focus construction without the nominative case marker. It is probable to postulate that Squliq Atayal is approaching the second stage of the universals of cleft-to-highlighting proposed in Harris
is no pause between the predicate *kun* and the nominative case marker *qu*. The discontinuity between the intonation lines of these two words not only lasts less than 0.1 second but also does not show up in the sound spectrum.

In addition, the GFOC *kun* bears the highest pitch and conveys a steep intonation movement. These two features definitely can arouse the hearer’s auditory perception and draw his/her attention in the process of communication. It is attested again from this example of Squliq Atayal that GFOC is always accompanied with prosodic prominence.

As mentioned above, the clausal subject of pseudo-cleft constructions is an open clause which lacks one argument. This missing argument in Squliq Atayal is the subject inside the open clause. This null subject must be coreferential with the nominal predicate (i.e. GFOC). To wit, pseudo-clefting in this language complies with the subject-only constraint: only subject can be pseudo-clefted: & Campbell (1995) and Harris (2001): that is, like copular verbs in other languages, the nominative case marker is undergoing grammaticalization to a focus particle.
In (290), the verb in the clausal subject is infixed with the AV marker -m-, which indicates that its subject must be the actor in terms of argument structure. Therefore, the only argument left in the clausal subject in question (i.e. *isu* ‘you’) can only function as the patient, not as the actor; otherwise, we will get a wrong meaning. That is to say, the missing argument is the actor as well as the subject of the clausal subject; the focalized nominal predicate *sami* ‘we/us’ plays these two roles. If the positions of these two free pronouns (*isu* ‘you’ and *sami* ‘we/us’) are exchanged, as shown in (291), it will follow that the argument staying in the clausal subject (*sami* ‘we/us’, this time) still can only be the patient and the object and that the missing argument is always interpreted as the actor and the subject of the event. The examples in (292) and (293) display a similar pattern. It is the missing arguments, with which the focalized nominal predicates are
co-referential, that serve as both patients and subjects. Based on the interpretation of these four sentences, it is obvious that grammatical subject plays a crucial role in a pseudo-cleft sentence.

In the same vein, LV and CV sentences obey the same constraint: no matter what thematic role the missing argument plays in the clausal subject, its syntactic role must be the subject inside the clausal subject. If we stick to the same meaning but pseudo-cleft non-subjects, then ungrammaticality will be born out.

\(294\) slaq qani qu’ qniq-an qulih watan.
  wet.field this Nom eat-LV fish Watan
  ‘Where Watan ate fish is this wet field.’ <E>

\(295\) *watan qu’ qniq-an qulih slaq qani.
  Watan Nom eat-LV fish wet.field this <E>

\(296\) *qulih qu’ qniq-an watan slaq qani.
  fish Nom eat-LV Watan wet.field this <E>

\(297\) qani qu’ s-qyanux=myan.
  this Nom CV-live=1PObl
  ‘What we live by is this.’ <E>

\(298\) *sami’ qu’ s-qyanux qani.
  1PObl Nom CV-live this <E>

The cases in (294), (295) and (296) are in LV. Given that they all have the same meaning as shown in (294), the sentence is grammatical only when the subject (i.e. the location here) slaq qani ‘this wet field’ is pseudo-clefted, as in (294); otherwise, it is ungrammatical no matter what argument, either the oblique actor watan or the object/patient qulih ‘fish’, is pseudo-clefted, as shown in (295) and (296). Likewise, sentences in CV, like (297) and (298), only permit the pseudo-clefting of the subject (i.e.
circumstance here) qani ‘this’ rather than other arguments like the actor sami’ ‘(by) us’.

So far, GFOC resembles GTOP in Squliq Atayal in terms of the subject-only constraint on the missing argument. However, they have the opposite reactions to resumption. While resumption is only employed in the possessor-possessed relationship of GTOP, it is employed in the subject position of GFOC even though it is optional and not common. As presented in the following sentence (299), a resumptive pronoun hiya’ ‘he’ fills in the missing subject inside the clausal subject:\footnote{Squliq Atayal is a language having internally-headed relative clauses (Liu 2004a, b). In (299), it seems to make no difference whether the free pronoun hiya’ is analyzed as a resumptive pronoun within an open clause or as a head noun within a head-internal relative clause.}

\begin{align*}
\text{(299) } &\text{sinsi qasa}_i\text{ qu’ wal p}nep \text{ qulih hiya’}. \\
&\text{teacher that Nom AV.went AV.angle fish 3SNeu} \\
&\text{‘The one who went angling is that teacher.’ <E>}
\end{align*}

\begin{align*}
\text{(300) } &\text{*sinsi qasa}_i\text{ qu’ b<n>hy-an=nya’}, \text{ mlikuy.} \\
&\text{teacher that Nom <Perf>hit-PV=3SObl man} \\
&\text{Intended for ‘The man who was hit by him is that teacher.’} \\
&\text{‘#The one by whom the man was hit is that teacher.’ <E>}
\end{align*}

\begin{align*}
\text{(301) } &\text{*yukan=su’}_i\text{ qu’ betunux yal syasin=nya’}. \\
&\text{Yukan=2SGen Nom beautiful very photo=3SGen} \\
&\text{Intended for ‘The man whose photo is very beautiful is your Yukan.’ <N>}
\end{align*}

If resumptive pronouns occur in non-subject positions, such as the oblique in (300) and the genitive in (301), these sentences will be ill-formed.

In addition, although it is commonly assumed that pseudo-cleft constructions mark focus, this assumption is not always correct. As stated in Givón (1990), “[pseudo-clefting] should not be considered per se a contrastive-focus device, since its nominal predicate may be either stressed or unstressed, and is contrastive only when stressed. (p.704)” The English examples are illustrated below:
(302) The one who will milk the goat is JOE. (Contrastive) (Givón 1993:178)

(303) The one who will milk the goat is Joe. (Non-contrastive) (Givón 1993:178)

(304) It is Joe that will milk the goat.

When the nominal predicate is stressed like (302), this pseudo-cleft construction conveys contrast, and it is the allosentence of (304).\textsuperscript{115} Lambrecht (1994) regards it as a real pseudo-cleft sentence with the identificational reading since the nominal predicate identifies the missing argument in the open clause. By contrast, the unstressed nominal predicate in (303) does not convey contrast; thus, sentence (303) is not the allosentence of (304), merely viewed as a topic-comment construction in Lambrecht (1994). Put simply, contrast is not an inherent property of pseudo-cleft constructions if we do not make a distinction between real/stressed pseudo-clefting and seeming/unstressed pseudo-clefting. Furthermore, although contrastive pseudo-clefting and clefting both carry contrastive focus, only clefting displays the exhaustiveness condition: the identified referent must be unique and the only one to fit the open argument position (Horn 1981, secondhand cited from Lambrecht (1994)).\textsuperscript{116} Pragmatically, clefting

\textsuperscript{115} Allosentences are a set of sentences which have the same propositional content but differ in the arrangement of information. It is a linguistic term used in Lambrecht (1994), also called “thematic variant(s)” in Huddleston (1988).

\textsuperscript{116} It is controversial among linguists why cleft constructions obey the exhaustiveness condition. The exhaustiveness condition may come from truth-functional effect (Atlas & Levinson 1981), uniqueness/maximality presupposition (Wedgwood, Pethö & Cann 2006), or conversational implicature. The neurolinguistic study of Drenhaus et al. (2011) showed that the exhaustiveness condition in question was not truth-functional and that its relation with pragmatics was still unknown. Avery D. Andrews (p. c.) pointed out that the exhaustiveness condition on cleft constructions might be a misnomer since cleft constructions eliminated the complementary subset; instead, cleft constructions should be exclusive. This can be illustrated from the examples below (from Ward, Birner & Huddleston (2002:1426)):

(i) That was hard work. What I need now is a long cool drink.

(ii) That was hard work. #'t's a long cool drink that I need now.

In the first example, although the pseudo-cleft construction makes the needs of the worker salient, it does not exclude other possibilities; in contrast, in the second example, the cleft construction excludes other possibilities, which actually do not exist in this mini discourse context, so the whole example is not felicitous.
always involves a strong assumption of the hearer’s contrary belief (Givón 1990).

Unlike English, pseudo-clefting is specialized as the contrastive-focus construction in Squilq Atayal. In the naturalistic data, all examples convey contrastive meaning, highlighting the preposed nominal against others. Take (305) for example. This sentence contrasts the person *watan* against other people. The equivalent of the non-contrastive pseudo-clefting is like the sentence in (306), where the headless relative clause *syon na’ sayun* ‘the one who is loved by Sayun’ is topicalized. Once the headless relative clause is topicalized, the nominal predicate *watan* is not in the sentence-initial position and its information does not carry contrast anymore.

(305) **watan** qu’ sy-on na’ sayun.

*Watan* Nom like-LV Obl Sayun

‘The one who is loved by Sayun is Watan.’ (Contrastive) <E>

(306) sy-on na’ sayun ga, **watan**.

like-LV Obl Sayun Top Watan

‘(As for the one who) is loved by Sayun, (he) is Watan.’ (Non-contrastive) <E>

Although Squilq Atayal has no cleft constructions like English, the exhaustive interpretation can be attained through pseudo-cleft constructions with the assistance of discourse, like the following sentence:

(307) kun qu’ m-wah yunaw m-lahung ngasal na’

1SNeu Nom AV-come AV.take.turn AV-support house Gen kaya qani.
deceased.mother this

‘The one who came to support this house of Mother in turns is me.’
or ‘It is I who came to support this house of Mother in turns.’ <N>

The above example is merely one part of a narrative, in which the speaker mentioned her stepfather’s devotions to the family and some of her relatives’ irresponsibility. Right
before the sentence (307), the narrative line reached her stepfather’s death and how shabby the old house was since that time; then, the speaker said this sentence to emphasize that she was the only one coming to take this responsibility. The exhaustive meaning comes from the discourse-pragmatic context, not from the pseudo-cleft construction itself. If discourse does not offer enough information for this reading, the modification of the adverb nanak ‘only’ can strengthen it:

(308) hiya’ nanak qu’ nyux s-m-ru’ balay la, sami’ ake, tayal
   3SNeu only   Nom  Aux.Prog AV-support really Par 1PNeu Par  Atayal
   qani   ga.
   this Par
   ‘The one who is really supporting us, these Atayal people, is only her.’
   or ‘It is she that is really supporting us, these Atayal people.’ <N>

The above sentence exists in a narrative, following a series of sentences introducing a woman who works hard to protect the Atayal culture. Because the speaker forgot to mention that this woman was the only person working on the preservation of the Atayal culture in their tribe, this sentence (308) was used to supplement this part. Notice that the occurrence of the adverb nanak ‘only’ shows that only pseudo-cleft constructions are not sufficient to express the exhaustive reading in this discourse. To wit, pseudo-cleft constructions in Squiliq Atayal per se have no requirement of the exhaustiveness condition, but with the assistance of discourse-pragmatic contexts or adverbs, the exhaustive reading can be derived.

8.2.2 Related Usage

This section falls into two parts, both of which are related with the extended usage of pseudo-clefting. The first one is about pseudo-clefted adverbials, and the second one is about interrogative constructions, most of which are expressed via pseudo-cleft sentences.
8.2.2.1 Adverbial Focus

In addition to the pseudo-clefting of argument, it is possible to pseudo-cleft temporal adverbials in Squiql Atayal, as pointed out in Hsiao’s (2004) study:

(309) hira’ qu’ m-ihiy yumin qu’ tali’. (Hsiao 2004:86)
    yesterday Nom AV-hit Yumin Nom Tali
    ‘It is yesterday that Tali hit Yumin.’

In the above sentence, the temporal adverbial hira’ ‘yesterday’ is focalized, and based on the meaning in the original study, it indeed displays contrast. As for adverbials of reason or location, they cannot be pseudo-clefted directly unless they become the subject via the CV or LV marking, as shown below:

(310) pya qu’ s-ngilis na’ laqi qani.
    candy Nom CV-cry Obl child this
    ‘What the child is crying for is candy.’ <E>

(311) *pya qu’ m-ngilis laqi qani.
    candy Nom AV-cry child this <E>

(312) ngasal qasa qu’ niq-an na’ yukan ngta’.
    house that Nom eat-LV Obl Yukan chicken
    ‘The place where Yukan eats chicken is that house.’ <E>

(313) *ngasal qasa qu’ m-aniq ngta’ yukan.
    house that Nom AV-eat chicken Yukan <E>

The difference between (310) and (311) lies in the voice marking and the grammatical function played by the word pya ‘candy’: in the former, the verb is marked with the CV prefix s- so the word pya ‘candy’ as the reason of the event is actually the subject, while
in the latter, the verb is marked with the AV prefix m- so the subject is laqi qani ‘this child’ (the actor of the event). The word pya ‘candy’ in (311) is just a pure adverbial, not an argument, and this sentence is ill-formed. It follows that an adverbial of reason cannot be pseudo-clefted in Squliq Atayal. Similarly, the location ngasal qasa ‘this house’ functions as the subject in (312) whereas it is a pure locative adverbial, not an argument, in (313). The ungrammaticality of (313) attests that locative adverbials cannot be pseudo-clefted.

Some adverbials seem to occur in pseudo-cleft constructions. These seemingly pseudo-clefted adverbials are ubiquitous in natural data, as exemplified below:

(314) si kta-y qu’ khmay qu’ squliq uzi la. (Rau 1992:238)
    if look-PV Nom many Nom people too Par
    ‘Lo and behold, the population increased.’

(315) ta qu’ hbun balay cyaqung lma ga. (Egerod 1969:299)
    look Nom Hbun really Cyaqung Par Par
    ‘Suddenly (they were at) Hbun Cyaqung.’

(316) nanu’ yasa qu’ syuk-un=nha’ zipun.
    Par thus Nom retaliate-PV=3PObl Japanese
    ‘Therefore, Japanese were retaliated against by them.’ <N>

(317) yasa qu’, m-zyup squ’ ngbang qani. (FLA)
    thus Nom AV-fall.into Loc cave this
    ‘As a result, you fell into the cave.’

In (314) and (315), the adverbials si kta and ta are grammaticalized from the verb kita ‘see’, but they cannot be treated as verbs anymore since the clausal subjects of these two examples are not the experiencers of perceptual actions. The examples (316) and (317) are related with yasa ‘thus, in this way’. No matter whether it co-occurs with with the word nanu’ ‘what/well’ or not, when it is in seemingly pseudo-cleft constructions, it
behaves like a transitional adverbial meaning ‘therefore’. It follows from these examples that these sentences are definitely not pseudo-cleft constructions because all of the adverbials are non-referring, as well as not contrasted with other referents. The above adverbials are associated with pseudo-cleft constructions only in their historical development. In the research of Lehmann (2008) and Traugott & Trousdale (2013), it is pointed out that there is a change usually observed in the development of pseudo-cleft constructions: specificational pragmatics associated with identifying referents of sets tends to give way to the pragmatics of indexing upcoming discourse.

8.2.2.2 Interrogative Constructions

Interrogative constructions have been an important issue of the studies on Atayalic languages, such as C. Chang (1995) and H. Chang (1997) on Paran Seediq, L. Huang (1996b) and L. Huang et al. (1999) on Mayrinax Atayal (the C’uli’ dialect), Lin (2005) on Jianshi Atayal, and Shih (2008) on Plngawan Atayal (the C’uli’ dialect). No matter what language or dialect, interrogative constructions in all Atayalic languages display the following characteristics:

(318)

(a) For nominal interrogatives as arguments, if they are subjects, they must be pseudo-clefted;
(b) If they are non-subjects, they must remain in their original positions (in-situ), still inside the predicates.
(c) For adverbial interrogatives, they can be either pseudo-clefted or in-situ.
(d) For verbal interrogatives, they must be in-situ.
(e) For adjectival interrogatives, they must be pseudo-clefted either alone or with the head nouns they modify.

Generally speaking, the distribution of interrogatives resembles that of their equivalent of non-interrogatives, except that interrogative subjects and adjectivals have to be pseudo-clefted. In the following paragraphs, these characteristics will be presented in
examples of Squiq Atayal. Most of linguistic facts are adopted from Lin (2005), which further benefits from other works mentioned above. In order to understand the use of pseudo-clefting on interrogatives, it is inevitable to repeat these discoveries.

As discussed in the a- and b-point of (318), if nominal interrogatives function as subjects, they have to be pseudo-clefted; otherwise, they can stay in their ordinary positions inside verb phrases. This fact tallies with the subject-only constraint since it is ungrammatical to pseudo-cleft non-subject arguments in this language:

(319) nanu’ qu’ sy-on na’ tali’. (Lin 2005:39)
    what Nom like-PV Obl Tali
    ‘What is liked by Tali?’

(320) *sy-on na’ tali’ qu’ nanu’. (Lin 2005:39)
    like-PV Obl Tali Nom what

(321) s-m-oya nanu’ qu’ tali’. (Lin 2005:40)
    AV-like what Nom Tali
    ‘What does Tali like?’

(322) *nanu’ qu’ s-m-oya qu’ tali’.
    what Nom AV-like Nom Tali <E>

It follows from (319) and (320) that nanu’ ‘what’ as both the patient and the subject of a PV sentence must be preposed in the focus part of pseudo-clefting, the violation of which results in ungrammaticality. Examples (321) and (322) buttress the b-point of (318). The interrogative object nanu’ ‘what’ must stay in its original position, and pseudo-clefting it violates the subject-only constraint on pseudo-cleft constructions. Although nominal interrogatives can be easily either pseudo-clefted or in-situ particularly with the alternation of multiple grammatical voices in this language, the

---

117 Some word-final glottal stops are missing in the data in Lin (2005) so they are added based on Egerod (1980, 1999).
frequency of their occurrence in natural texts is not fifty-fifty; instead, the pseudo-cleft type in (319) plays an absolutely dominant role.

Like most of other Formosan languages, Squliq Atayal has few types of adverbs. Most of the adverbs in English (especially manner adverbs) correspond to verbs in this language. The adverbial interrogatives in the c-point of (318) have only one member in Squliq Atayal: *knwan ‘when’, the interrogative of temporal adverbials. Its distribution is very flexible, as shown below:

(323) (knwan) p-m-uya (*knwan) pagay (knwan) qu’ tali’ (knwan).
  when Ft-AV-grow when rice when Nom Tali when ‘When will Tali grow rice?’ (Lin 2005:43)

(324) knwan qu’ p-m-uya pagay qu’ tali’. (Lin 2005:47)
  when Nom Ft-AV-grow rice Nom Tali ‘When will Tali grow rice?’

Example (323) demonstrates that the interrogative adverbial *knwan ‘when’ can be preverbal, or between the object and the nominative case marker *qu’, or after the subject. The only place where it cannot occur is the slot between verbs and their objects. Its distribution is exactly the same as that of temporal adverbials. In addition, it can be pseudo-clefted like (324).119

The interrogative *hmswa ‘why/how’ is undoubtedly a verb since it has different

[118] It is common in Formosan languages that manner adverbial modifiers are verbs. Chang (2006) and Holmer (2006) analyze these manner adverbials, in Kavalan and Seediq respectively, as functional heads in syntactic configurations.

[119] Lin (2005) pointed out that there was topicalization of the adverbial interrogative in Squliq Atayal and that it does not exist in the sister language Seediq (Cf. H. Henry Yungli Chang 1997), as exemplified below:

(i) knwan ga, p-m-uya pagay qu’ tali’. (Lin 2005:48)
  when Top Ft-AV-grow rice Nom Tali ‘When will Tali grow rice?’

However, according to my informants, topicalization of interrogatives is unacceptable unless the interrogative word is interpreted as an indefinite pronoun. In addition, it is not attested in our naturalistic data, either.
conjugations like *swa’un* (PV), *swa’an* (LV), etc. Moreover, it can function as predicate in a simple sentence, as illustrated in (325):

(325) h-m-swa’ qu’ qutux, iy, s-on=nha’.  (Egerod 1999:301)  
AV-why Nom one Fil say-LV=3PObl  
‘How is that other story they tell?’

(326) h-m-swa’ (*qu’) m-usa mnka qu’ tali’. (Lin 2005:57)  
AV-why Nom AV-go Taipei Nom Tali  
‘Why does Tali go to Taipei?’

Usually, the adverbial interrogative *hmswa’* ‘why/how’ comes with other verbs to form a serial verb construction like (326). Since Squliq Atayal has no real manner adverbs but express the meaning by means of verbs, the verbal interrogative is identical with these verbs. Notice that the nominative case marker *qu’* cannot be inserted in between the two verbs *hmswa’* ‘why/how’ and *musa* ‘go’. This proves that verbal interrogatives cannot be pseudo-clefted.

Adjectival interrogatives are analyzed as verbal interrogatives in L. Huang (1996b), L. Huang et al. (1999), and Shi (2008), while they are separated as an independent type in Lin (2005) because of their peculiar syntactic properties. This gap comes from the syntactic distributions of adjectival interrogatives. Adjectival interrogatives are modifiers of noun phrases, so they can be pseudo-clefted with their modifiees as long as their modifiees are subject:

(327) *qenu hozil* qu’ sy-on=su’.  (Lin 2005:59)  
which dog Nom like-PV=2SObl  
‘Which dog is liked by you?’
In the above two sentences, adjectival interrogatives (qenu ‘which’ and pira ‘how many’) and their modifiees (hozil ‘dog’ and ttu ‘trap’) are pseudo-clefted together. Furthermore, adjectival interrogatives can be pseudo-clefted alone, leaving their modifiees behind in the subject position:

(329) **qenu** qu’ sy-on=su** hozil. (Lin 2005:59)**

which Nom like-PV=2SObl dog
‘Which is the dog liked by you?’

Compared with (327), the adjectival interrogative qenu ‘which’ is separated with its head noun hozil ‘dog’ in (329) by a string of words. Notice that the nominative case marker qu’ appears in between the adjectival interrogative qenu ‘which’ and the verb syon ‘like’, which is impossible with verbal interrogatives as shown in (326). It will violate the subject-only constraint if it is claimed that only the modifier of subject is pseudo-clefted. In order to account for these two patterns, Lin (2005) proposes that these adjectival interrogatives can function as either heads or modifiers. When the whole noun phrases are pseudo-clefted as in (327) and (328), adjectival interrogatives are modifiers; on the other hand, when only adjectival interrogatives are pseudo-clefted as in (329), they are heads, (to my interpretation) the predicates of the whole sentences. Although Lin’s (2005) proposal is generally on the right track, its explanatory power is weak partly because the same set of adjectival interrogatives are treated as two different sets with identical members, and partly because it still cannot explain why adjectival interrogatives as modifiers cannot remain inside the original subject position. In other words, there is no consistent analysis of these two patterns. A more challenging example comes from Lin’s (2005) own data:
(330) **pira** qu’ **squiliq** m-m-usa q-m-alup. (Lin 2005:62)

how many Nom people Red-AV-go AV-hunt

‘How many people will go hunting?’

If *pira* ‘how many’ is a modifier, then we cannot explain why there is a nominative case marker *qu*’ intervening between *pira* ‘how many’ and its modifiee *squiliq* ‘people’. On the other hand, if *pira* ‘how many’ is a head, then why is the subject preposed in the second half of this sentence? Although it is possible to have two focalized elements in natural languages, it is irrational to claim that there is a language which splits a noun phrase into two parts and then pseudo-clefts them respectively.

Contrary to the above dichotomous analysis, we can get a more consistent analysis if the construction of internally-headed relative clauses (IHRCs) is introduced.¹²⁰ It has been attested in Liu (2004a, b) that Squliq Atayal has IHRCs, where the head nouns modified by relative clauses exist inside relative clauses themselves, and that the whole IHRCs behave like complex noun phrases. To be more specific, it is claimed in the present thesis that adjectival interrogatives only function as predicates.¹²¹ There is no

---

¹²⁰ Whether the IHRC analysis fits into the facts of Seediq, the sister language of Atayal, needs future research. Unlike Squliq Atayal, Seediq permits the pseudo-clefting of adjectival interrogatives even when their head nouns are obliques, as illustrated below:

(i) kenu (ka) [hma-un na qhuni swai rseno]. (C. Chang 1995:35)

which Nom plant-PV Obl tree younger.sibling male

‘Which is the tree younger brother planted?’

If the whole subject is analyzed as an IHRC with the internal head noun *qhuni* ‘tree’, then we have to claim that Seediq allows obliques to be relativized in IHRCs. This claim will bring some controversies. First, IHRCs are still unattested in Seediq. Second, it contradicts the fact that only subject can be relativized in externally-headed relative clauses (EHRCs) in this language (H. Henry Yungli Chang 1997, Chang 2000).

¹²¹ This claim may further arouse the doubt about whether all modifiers of noun phrases (or adjectivals) in Squliq Atayal are predicates in nature. The answer is negative. According to Liu (2004b), there are two kinds of nominal modifications: attributive and predicative. The former is substantiated in *na*-constructions, while the latter is unmarked with any linker. The difference between attributive modification and predicative modification is verifiable through the semantic contradiction test:

(i) tali’ ga, wagiq na’ knsingut. (Liu 2004b:69, fn.22)

Tali Top tall Lin midget

‘Speaking of Tali, (he) is a tall midget.’

(ii) ??tali’ ga, wagiq Ø knsingut. (Liu 2004b:69, fn.22)

Tali Top tall Ø midget

260
distinction between “modifier” and “head”. Both of them are actually the predicates in IHRCs. Then, our analyses of (327), (329) and (330) are like the following:

(331) [qenu hozil] qu’ sy-on=su’. (=327)
   which dog Nom like-PV=2SObl
   ‘What is liked by you is which dog?’ (lit.) <E>

(332) qenu qu’ [sy-on=su’ hozil]. (=329)
   which Nom like-PV=2SObl dog
   ‘The dog liked by you is which?’ (lit.) <E>

(333) pira qu’ [squliq m-m-usa q-m-alup]. (=330)
   how-many Nom people Red-AV-go AV-hunt
   ‘The people going hunting are how many?’ (lit.) <E>

IHRCs are all bracketed in the above sentences. In (331), qenu ‘which’ is the predicate of hozil ‘dog’, and they constitute an IHRC. The whole IHRC qenu hozil ‘which dog’ is pseudo-clefted. Because IHRCs are like complex noun phrases, they definitely can function as subjects and be further pseudo-clefted. In (332), qenu ‘which’ is also a predicate, but its subject is the IHRC syon-su’hozil ‘the dog liked by you’. Note that the head noun of IHRCs can only be a subject (Liu 2004a, b). Therefore, the interpretation of the whole IHRC is never ‘you liked by the dog’. The interrogative qenu ‘which’ still remains located in the main predicate. The structure of (333) is like that of (332). The only difference is that the head noun squliq ‘people’ is preposed inside the IHRC. As

‘Speaking of Tali, (he) is a midget, who is tall.’

The referent of knsingut ‘midget’ is short, which runs afoul with the property of wagiq ‘tall’. The only way to make wagiq ‘tall’ modify knsingut ‘midget’ is placing the adjectival in an attributive modification which makes the property gradable. Therefore, the example (i) gives us the interpretation that the midget is relatively tall compared with other midgets. On the contrary, the example (ii) is semantically contradictory since a midget cannot be both short and tall in nature. Notice that adjectival interrogatives presented here are not in na’-constructions at all. Our analysis of the adjectival interrogatives dovetails Liu’s (2004b) distinction on these two kinds of modifications: that is, all adjectival interrogatives are in predicative modification relationship. In addition, it is highly possible that na’-constructions are real adjunction in terms of syntactic configuration.
proven in other natural languages, embedded clauses can have their own focalized element. It is unproblematic that there is another element focalized in an IHRC. In fact, indirect questions in Squiliq Atayal do have preposed interrogatives. For instance, the adjectival interrogative qenu ‘which’ is preposed even inside the embedded clause qenu babaw-nya ‘in which way their future is’.

\[(334) \text{baq-aw=ta' } m-\text{usa } m-\text{ha } [\text{qenu babaw=nya'}] \text{ la } \text{ pi.}^{122} \]
\[\text{know-PV Proj=1PObl AV-go AV-say which future=3SGen Par Par} \]
\[\text{‘We do not know in which way their future is.’} \text{ <N>} \]

The proposal in the present thesis not only bridges the gap between the classification in L. Huang (1996b), L. Huang et al. (1999), and Shi (2008) and that in Lin (2005), but also strengthens their respective classifications. Verbal interrogatives and adjectival interrogatives are similar in that they both function as predicate in nature, while they are distinctive from each other in the means they employ to obtain the function of modification: verbal interrogatives do it by “compounding” themselves with their modifiees where adjectival interrogatives by “relativizing” their modifiees.

All in all, this section reviews the five characteristics of interrogative constructions in Squiliq Atayal. No matter whether interrogatives are pseudo-clefted or in-situ, they can never occur in subject position. Moreover, it is demonstrated that under the IHRC analysis proposed in the present thesis, the peculiarities of adjectival interrogatives can be consistently explained.

### 8.3 Theoretical Issues

This section discusses the theoretical analysis of syntactic structure of

---

122 Although there is no negative word in this sentence, the free translation given by the informant does show the negative meaning, which comes from the expression bagaw-ta’ ‘we do not know’ in Egerod (1999). In addition, it happens often in natural texts that informants use the third person singular pronoun -nya’ to refer to plural people.
pseudo-clefting and then of the discourse functions in i-structure it corresponds to. The end of this section will discuss the general mapping of GFOCs and the issues on the lexical properties of interrogative nominals as well as the hypothetical development of pseudo-clefted transitional adverbials.

### 8.3.1 Syntactic Structure

Pseudo-clefting is the only GFOC construction in Squiliq Atayal. Like *wh*-cleft sentences in English, pseudo-cleft sentences in Squiliq Atayal have a nominal predicate and a headless relative clause as the grammatical subject of the matrix clause. Like other verbless clauses in this language, the nominal predicate in a pseudo-cleft sentence is not inflected with predicative morphology such as tense, aspect, mood and subject-verb agreement, which differs from some languages with verbless clauses in which nominal predicates morphologically behave like verbal ones (Cf. Curnow 2000).

(335) *sami’ qu’ k-m-ut isu’ Ø*.

1PNeu Nom AV-kill 2SNeu Ø

‘The one who killed you is us.’ <N>

The headless relative clause is a finite clause, not a non-finite one, because the embedded predicate inside it may contain auxiliaries of different tenses, as shown in (336) and (337).

(336) *sami’ qu’ wal k-m-ut watan.*

1PNeu Nom Aux.Pt AV-kill Watan

‘The one who killed Watan is us.’ <E>

(337) *sami’ qu’ musa k-m-ut watan.*

1PNeu Nom Aux.Ft AV-kill Watan

‘The one who is going to kill Watan is us.’ <E>
The example in (336) has a headless relative clause with the auxiliary *wal*, which indicates that the event happened in the past, while the headless relative clause in the example in (337) contains the auxiliary *musa*, which indicates that the event will happen soon in the nearby future. Therefore, although the headless relative clause serves as the grammatical subject in a matrix clause, it is not a noun phrase in nature, still being a finite clause, since noun phrases do not have tense marking or auxiliaries. In addition, any verb can serve as the predicate inside the headless relative clause, which means that the verb is not lexically nominalized since not all verb have a lexically nominalized counterpart in the lexicon (Chomsky 1970).

As for the matrix clause in a pseudo-cleft sentence, it is a verbless clause which aims to identify the predicative nominal.\(^{123}\) In LFG, there are two ways to analyze verbless clauses. One is the single-tier approach, in which a non-verbal predicate carries PRED in a single-tier f-structure, and the other is the double-tier approach, in which there is an invisible copular verb that subcategorizes the non-verbal predicate as its argument. The former originates from Butt et al. (1999) and the latter from Andrews (1982b) and Grimshaw (1982). Attia (2008) and Butt et al. (1999) propose that the single-tier analysis is universally held for all kinds of verbless clauses. In contrast, some studies, like Dalrymple, Dyvik & King (2004), Falk (2004), Laczkó (2012), and Nordlinger & Sadler (2007), believe that for the validity of these two analyses, there can be variation across languages and also across constructions within the same language. Thus, which of the two analyses is appropriate differs from case to case. In order to capture the syntactic structure of pseudo-clefting in Squilq Atayal, we have to investigate that of ordinary identificational verbless clauses first. Take the example of

\(^{123}\) According to Avgustinova & Uszkoreit (2003) and Avgustinova (2006), the relations between a subject and a non-verbal predicate (mediated by a copular verb) can be classified into several types: attributive, classificational, equative, specificational, existential, locative and possessive. Whether there is a copula verb varies from language to language, and even from relation to relation within a language. In other words, copula constructions of these relations may not have a copular verb, and those without a copular verb are verbless clauses.
(338), a verbless clause with a predicative nominal. The grammatical subject is *sayun* and the predicative nominal *yaya ‘-nya’ ‘his mother’* serves to indicate the identity of the grammatical subject. The single-tier analysis and the double-tier analysis of this example are presented in (339) and (340) respectively.\(^\text{124}\)

\[(338) \text{yaya‘-nya’ qu’ sayun.} \]
\[
\text{mother=3SGen Nom Sayun} \\
\text{‘Sayun is his mother.’ } <\text{E}> \\
\]

\[(339) \text{The single-tier analysis} \]

\[
\begin{array}{c}
\text{PRED} \quad \text{'mother (POSS, SUBJ)'} \\
\text{SUBJ} \quad [\text{PRED} \quad \text{'Sayun'}] \\
\quad [\text{PRED} \quad \text{'his'}] \\
\text{POSS} \quad \text{CASE Gen} \\
\quad \text{PERS 3} \\
\quad \text{NUM Sg} \\
\end{array}
\]

\[(340) \text{The double-tier analysis} \]

\[
\begin{array}{c}
\text{PRED} \quad \text{'be (SUBJ, PREDLINK)'} \\
\text{SUBJ} \quad [\text{PRED} \quad \text{'Sayun'}] \\
\quad [\text{PRED} \quad \text{'mother (POSS)'}] \\
\quad [\text{PRED} \quad \text{'his'}] \\
\text{PREDLINK} \quad \text{POSS} \\
\quad \text{CASE Gen} \\
\quad \text{PERS 3} \\
\quad \text{NUM Sg} \\
\end{array}
\]

In (339), the predicative nominal *yaya ‘-nya’ ‘his mother’* itself is the predicate of the whole clause—that is why it is called the single-tier analysis—and it subcategorizes two arguments POSS (possessor) and SUBJ, the former of which is the genitive pronoun *-nya*.

\(^{124}\) Thanks to the anonymous examiner’s reminding, there is a third approach which is also a double-tier analysis involving XCOMP.

265
‘his’ within the predicative nominal and the latter of which is sayun the grammatical subject of the whole verbless clause. In contrast, yaya ‘-nya’ ‘his mother’ in (340) occurs in the embedded matrix of PREDLINK so that it is called the double-tier analysis. Strictly speaking, the predicative nominal is not a predicate of the matrix clause but an argument subcategorized by the predicate of the matrix clause. The real predicate is an invisible copular verb which does not appear in c-structure but contributes its annotations to f-structure, symbolized as ε (epsilon) (Laczkó 2012).

In the present thesis, it is assumed that the double-tier analysis is more appropriate to identificational verbless clauses in Squliq Atayal. With everything being equal, the single-tier analysis has at least two drawbacks in the present case. First, although the single-tier analysis perfectly accounts for languages with non-verbal predicates inflected with verbal morphology, Squliq Atayal does not belong to this kind of languages, as exemplified in (341). The predicative nominal cannot be prefixed with the AV marker m- in comparison with (338).

(341) *m-yaya’=nya’ qu’ sayun.
   AV-mother=3SGen Nom Sayun <E>

Second, as mentioned in Nordlinger & Sadler (2007), it remains an open question whether non-verbal predicates can subcategorize both SUBJ and POSS as in the single-tier analysis. Empirically speaking, the evidence from Squliq Atayal does not support the single-tier analysis. The ungrammatical example in (342) and its f-structure in (343) prove that a nominal does not subcategorize both SUBJ and POSS in Squliq Atayal since if it did, the subcategorized SUBJ should be able to occur everywhere it exists, as shown in the f-structure in (343), which contains the embedded matrix (marked in blue) identical with that in (339).
(342) *m-aniq qulih qu’ yaya’=nya’ qu’ sayun.
    AV-eat fish Nom mother=3SGen Nom Sayun
    Intended for ‘Sayun being his mother is eating fish.’ <E>

(343)

```
  PRED  'eat (SUBJ, OBJ)'
    [PRED  'mother (POSS, SUBJ)']
    SUBJ  [PRED  'Sayun']
    [PRED  'his']
    POSS  CASE  Gen
    [PERS  3]
    [NUM  Sg]
    OBJ   [PRED  'fish']
```

How to exclude (342) and (343) from the single-tier analysis needs an extra mechanism
differentiating nominals with subcategorized SUBJ from those without it, which is not as
straightforward as the double-tier analysis. The example in (342) can easily be excluded
from the double-tier analysis since the predicate maniq ‘eat’ does not subcategorize a
verbless clause.

The above two pieces of evidence prove that with regard to the syntactic structure
of identificational verbless clauses in Squilq Atayal, the double-tier analysis is more
appropriate than the single-tier analysis. Then, we may turn to our core issue in this
section, the syntactic structure of pseudo-clefting in Squilq Atayal. It is a verbless clause
with a predicative nominal and, unlike (338), a headless relative clause. Take (335) for
example. Adapting the ideas from previous LFG studies on headless relative clauses,
such as Camilleri (2011), Chatsiou (2010) and Lipps (2011), the present thesis proposes
that pseudo-clefting is constituted by the following phrase structure rules:

```
(344)  S → DP  DP
       (↑ PREDLINK) = ↓
       (↑ PREDLINK) = (↑ GOF2)
       (↑ PRED) = 'PRO'
       (↑ SUBJ PRED) = 'PRO'
       (↑ SUBJ PRON) = Free
```
The phrase structure rule in (344) indicates that this is a clause containing a nominal predicate (i.e. the first DP) whose grammatical function is PREDLINK, as shown in the annotation \((↑ \text{PREDLINK}) = \downarrow\), which functionally controls GFOC, with the same content between them, as required by the constraint \((↑ \text{PREDLINK}) = (↑ \text{GOF2}). In addition, the second DP functions as the grammatical subject whose predicate is a free null pronoun as indicated in \((↑ \text{SUBJ PRED}) = \text{'PRO'}\) and \((↑ \text{SUBJ PRON TYPE}) = \text{Free}. The second phrase structure rule in (345) states that a DP immediately dominates an IP whose grammatical function is an adjunct, as shown in \(\downarrow \in (↑ \text{ADJ})\), and a relative clause, as shown in \((↑ \text{ADJ TYPE}) = \text{Rel}. Furthermore, within this adjunct is an operator whose content is identical with the null pronoun the adjunct modifies, as required by the constraint \((↑ \text{ADJ OPER}) = ( (ADJ ↑) \text{PRED 'PRO'})\), and with the grammatical subject inside the adjunct, as required by the constraint \((↑ \text{ADJ OPER}) = (↑ \text{ADJ SUBJ}). The phrase structure rule in (346) manifests IP’s immediate dominance of VP, and the one in (347) states that there is an object inside the VP, as indicated in \((↑ \text{OBJ}) = \downarrow\).

The above phrase structure rules constitute the mapping illustrated in (348) below:
(348)
e-structure

\[
\begin{align*}
S & \quad \text{DP} \quad \text{DP} \\
& \quad \text{sam}i' \quad \text{IP} \\
& \quad \text{us} \quad \text{VP} \\
& \quad \text{V} \quad \text{DP} \\
& \quad \text{k-m-ut} \quad \text{isu'} \\
& \quad \text{AV-kill} \quad \text{you}
\end{align*}
\]

f-structure

\[
\begin{align*}
PRED & \quad '\text{be} \langle \text{SUBJ}, \text{PREDLINK} \rangle' \\
PRED & \quad '\text{PRO}' \\
PRONTYPE & \quad \text{Free} \\
\text{PRED} & \quad '\text{kill} \langle \text{SUBJ}, \text{OBJ} \rangle' \\
\text{OPER} & \quad [ ] \\
\text{SUBJ} & \quad [ ] \\
\text{OBJ} & \quad \{ \}
\end{align*}
\]

\[
\begin{align*}
PRED & \quad '\text{us}' \\
\text{CASE} & \quad \text{Neu} \\
\text{PERS} & \quad 1 \\
\text{NUM} & \quad \text{Pl} \\
\text{GOF2} & \quad [ ]
\end{align*}
\]
As mentioned in Section 8.2.2.1, temporal adverbials can be pseudo-clefted, as the example in (309) repeated below. It can be analyzed as a headless relative clause containing an operator coreferential with the IP-adjunct inside the relative clause, as stated in the phrase structure rule (349), which resembles (345) except the constraint $(↑ ADJ OPER) = (↑ ADJ ADJ)$. To lump (345) and (349) together, this constraint can be modified as (350), which requires that the operator should be coreferential with either the subject or the IP-adjunct inside the free relative clause. (352). Moreover, the phrase structure rule in (351) replaces the one in (346) so that there is a subject inside the free relative clause.

(309) hira’ qu’ m-ihiy yumin qu’ tali’. (Hsiao 2004:86)
   yesterday Nom AV-hit Yumin Nom Tali
   ‘It is yesterday that Tali hit Yumin.’

(349)  DP $\Rightarrow$ IP
       $\downarrow \in (↑ ADJ)$
       $(↑ ADJ TYPE) = Rel$
       $(↑ ADJ OPER) = (ADJ ↑) PRED 'PRO')$
       $(↑ ADJ OPER) = (↑ ADJ ADJ)$

(350)  $(↑ ADJ OPER) = (↑ ADJ \{SUBJ|ADJ\})$

(351)  IP $\Rightarrow$ VP
       DP
       $(↑ SUBJ) = \downarrow$

The mapping between c-structure and f-structure is illustrated below:
8.3.2 Mapping to DFs

What discourse functions do GFOCs correspond to? In order to get the answer, it is necessary to look into quantitative evidence first. The formation of GFOC is irrelevant to voice alternation and valence. Therefore, all tokens are taken into consideration instead of merely investigating the statistical data presented in Chapter 5, which only include those from transitive clauses. Table 8-1 below demonstrates the information
status of all GFOCs in terms of givenness and aboutness.

<table>
<thead>
<tr>
<th>[GIV]</th>
<th>[ABT]</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>17 (56.7)</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>2 (6.7)</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>5 (16.7)</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>6 (20.0)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

All the four categories have tokens, but the majority of tokens are distributed in [+GIV, +ABT] (56.7%), which is similar to GTOPs in that the major information they convey is given and important information. If only looking at givenness, we can find out that the ration of given information (63.4%) and new information (36.7%) is around 1.7 to 1, much smaller than that in GTOP-SUBJs (2.8:1) (Cf. Section 7.4.2.1). Apparently, new information plays a more important role in GFOCs. Even so, Table 8-1 shows that the information highlighted by pseudo-cleft sentences is not only new information but also given information in discourse-pragmatics. Although we will continue making use of the term “grammaticalized focus” for the pseudo-clefted element, it is merely a structural notion, not guaranteeing that the information is new in discourse-pragmatics.

Moreover, givenness and aboutness do not suffice to pinpoint the discourse functions conveyed by GFOCs in Squiliq Atayal. As discussed in Section 8.2.1, pseudo-clefting in this language is characteristic of contrast, i.e. [+CONT] in i-structure. Non-contrastive pseudo-cleft sentences only exist when the headless relative clause is topicalized. In the first category of Table 8-1, the discourse function of these tokens is contrastive topic (CSTOP), which is important prominent given information with
contrastive meaning. The occurrence of grammaticalized contrastive topics does not have to be in pairs. On the contrary, most of the time they occur in one clause, contrastive with other unmentioned members in the discourse universe, as demonstrated in the case of pseudo-cleft the third person unbound pronoun hiya’ ‘he/him’ in (308) (Cf. Section 8.2.1).

If the GFOC is a piece of given information which does not recur in subsequent discourse, that is, the second category in Table 8-1, it corresponds to contrastive theme (CSTHM) in i-structure. As shown in Table 8-1, there are only two tokens in our naturalistic data, and one of them is displayed in (353).

(353) r’tung qu’ nyux=myan kt-an aw baq kruyux.
     minute.hand Nom Aux.Prog=1PObl see-LV Par AV.know hour.hand
     ‘What we were thinking of as the hour hand was the minute hand.’ <N>

This sentence is excerpted from a ghost story, in the beginning of which the addresser and her husband mistook the minute hand for the hour hand on a clock dial so they went out on a dark deep night, much earlier than they had expected. Both the hour hand and the minute hand have been mentioned prior to this sentence. Thus, the pseudo-clefted nominal r’tung ‘minute hand’ is a piece of given information, but it does not recur anymore since the information flow moves onward to what she met on the way. The referent of r’tung ‘minute hand’ is highlighted in pseudo-clefting to make a contrast with other hands on the clock dial.

It is undoubted that the pseudo-cleft construction is able to make new information a GFOC even though the proportion is lower than we may expect. The example in (354) is one of the tokens in the third category of Table 8-1. It is excerpted from a spoken history of the Slaq tribe. The pseudo-clefted noun phrase nanak qutux yumin sugi ‘only Yumin Sugi’s family’ is not mentioned in previous discourse, and it is contrastive with other families immigrating to the locus later. Hence the GFOC here corresponds to
contrastive focus (CSFOC) in i-structure.

(354) nanak qutux yumin sugi <n>aki’ qani.
only one Yumin Sugi <Perf>live this
‘The one who had lived in this place was only Yumin Sugi’s family.’ <N>

As for the fourth category in Table 8-1, all of its tokens are interrogative sentences with pseudo-clefted interrogative words, as exemplified in (354). The interrogative word *ima’ ‘who’ is pseudo-clefted, and its referent is an information gap (when there are no conversational implicatures) introduced into the information flow for the first time, usually to highlight that there exists some member of the universe fitting into the open proposition, contrastive with other members of the universe. In subsequent discourse, the information gap brought by interrogative words will be filled so it is not repeated anymore. Thus, the tokens of the fourth category correspond to interrogative focus (QFOC) in i-structure.

(355) ima’ m-usa m-ita isu’.
who AV-go AV-see 2SNeu
‘Who went to see you?’ <N>

The abovementioned four discourse functions are not derived from any discourse function via the conversion function $Ϙ$ in i-structure, belonging to the independent type of one-to-many mapping. Although the majority is CSTOP, the other three categories do not form gradience from prototype to non-prototype, as verified in the fact that the discourse function of the opposite features in givenness and aboutness, QFOC, is the second largest group while the closest discourse function to CSTOP, that is CSTHM, has the fewest tokens. At least synchronically, we cannot tell how the relationship between these four discourse functions is affected by the conversion function $Ϙ$, even though it could have had influence diachronically.
8.3.3 General Mapping of GFOC

A pseudo-cleft sentence, which consists of a predicative nominal and a headless relative clause as the grammatical subject, is a marked construction for GFOC. Its phrase structure rule is shown in (356). This phrase structure rule contains two DPs. The first DP is a predicative nominal carrying the grammatical function PREDLINK in f-structure and either of the four discourse functions, contrastive topic, contrastive theme, contrastive focus and interrogative focus, as marked in \{(↑_i CSTOP) = ↓, (↑_i CSTHM) = ↓, (↑_i CSFOC) = ↓, (↑_i QFOC) = ↓\}. The symbol preceding each discourse function indicates the proportion of that discourse function: ♣ stands for 50% < & ≤ 70%, ● stands for 0% < & ≤ 10%, and ♥ stands for 10% < & ≤ 30% (Cf. Chapter 5). The values of GIV and ABT depend upon what discourse function the nominal predicate corresponds, as displayed in the four condition constraints, so they are unspecified and not annotated. However, all of these four discourse functions must be prominent and contrastive, as marked in (↓_i PROM) =c + and (↓_i CONT) =c + respectively. The second DP is the real grammatical subject, as indicated in (↑ SUBJ) = ↓.

\[
\begin{align*}
(356) \quad S & \rightarrow \quad \text{DP} & \quad \text{DP} \\
& & (↑ PREDLINK) = ↓ & (↑ SUBJ) = ↓ \\
& \left\{ \\
& \quad (↑_i CSTOP) = ↓ \\
& \quad (↑_i CSTHM) = ↓ \\
& \quad (↑_i CSFOC) = ↓ \\
& \quad (↑_i QFOC) = ↓ \\
& \right\} \\
& & (↑ CSTOP) = ↓ \rightarrow (↓_i GIV) =c + \land (↓_i ABT) =c + \\
& & (↑ CSTHM) = ↓ \rightarrow (↓_i GIV) =c + \land (↓_i ABT) =c - \\
& & (↑ CSFOC) = ↓ \rightarrow (↓_i GIV) =c - \land (↓_i ABT) =c + \\
& & (↑ QFOC) = ↓ \rightarrow (↓_i GIV) =c - \land (↓_i ABT) =c - \\
& & (↓_i PROM) =c + \\
& & (↓_i CONT) =c +
\end{align*}
\]
The correspondences between c-structure, f-structure and i-structure in the pseudo-cleft construction are demonstrated in (357). The discourse functions of independent one-to-many mapping are placed in the curly brackets. The mapping between c-structure and i-structure is indicated by the pink line, and the mapping between c-structure and f-structure is indicated by the green line. As for the mapping between f-structure and i-structure, indicated by the purple line, it is between the above mentioned discourse functions and GOF2, not PREDLINK since ordinary predicative nominals do not correspond to these discourse functions. The equation between PREDLINK and GOF2 is the distinguishing characteristic of pseudo-clefting.
(357)
c-structure

f-structure

i-structure
8.3.3.1 Contrast Position

The mapping in (357) is similar to the mapping of GTOP in that the mapping from either c-structure or f-structure to i-structure is one-to-many; that is to say, several discourse functions are mapped from one c-structural position and from one grammatical function. Apparently, it manifests again the fact that Squliq Atayal is an f-structural language in which the mapping between c-structure and f-structure is more consistent. However, the four discourse functions have something in common: they are all contrastive prominent information. In Squliq Atayal, it is more accurate to regard the nominal predicate of pseudo-cleft sentences as a contrast position than to treat it as a focus position although the linguistic term “grammaticalized focus (GFOC)” has been used throughout the present thesis. The contrast position is not unusual in linguistic typology. Some languages, like Gawwada (African language), have a specific syntactic position for contrastive information without any distinction on its givenness (Tosco 2010). Broadly speaking, it is rational to claim that GTOP marks grammaticalized non-contrastive prominent given information while GFOC marks grammaticalized contrastive prominent information (Cf. Section 7.2).

If we take the perspective of the linearity of information, we can perceive that this language tends to place prominent information, whether contrastive or not, in the beginning of a proposition. Combining what we discovered in unmarked correspondences, we can get the following ranking of i-structural constraints, where the Prominence-First Constraint ranks over the Aboutness-Last Constraint and the Aboutness-Last Constraint further ranks over the Given-First Constraint:

(358) \textbf{Prominence-First} \gg \textbf{Aboutness-Last} \gg \textbf{Given-First}

The ranking in (358) presents the ideal arrangement of information in this language: that
is, prominent information must precede nonprominent information, and among nonprominent information, the information important to subsequent discourse must follow the unimportant one. Nevertheless, it has to be re-emphasized that i-structural ranking has its limitation. First of all, it only works on i-structure, especially the arrangement of information, and factors in other modules are out of its control. Take the relative position between GTOP and GFOC for example. GTOP is a CP-adjunct in c-structure while GFOC is a nominal predicate in c-structure. CP-adjuncts precede the whole IP/S, so they definitely precede nominal predicates. That is why GTOPs have to precede GFOCs, as illustrated in the comparison between (359) and (360):

(359) hira’ ga, watan qu’ wal q-m-alup.
yesterday Top Watan Nom Aux.Pt AV-hunt
‘Yesterday, the one who went hunting was Watan.’ <E>

(360) *watan, hira’ ga, qu’ wal q-m-alup.
Watan yesterday Top Nom Aux.Pt AV-hunt <E>

The former sentence is grammatical while the latter sentence is not since the CP-adjunct hira’ ‘yesterday’ (i.e. GTOP) must precede the nominal predicate watan (i.e. GFOC). The grammaticality has nothing to do with the ranking of i-structural constraints.

Second, as discussed in Chapter 4, what the ranking represents is merely an ideal pattern. The real discourse fact is that information may be affected by the conversion function $Q$, which brings tokens deviating from this ideal pattern. In other words, the ranking of i-structural constraints only represents the general and ideal information arrangement of a language and it does not have the super power over all modules.

8.3.3.2 Lexical Properties of Interrogative Nominals

As presented in previous sections, QFOC plays an important role in the pseudo-cleft
construction. In fact, interrogative words, such as *ima* ‘who’ and *nanu* ‘what’, are lexically marked with their discourse function QFOC, as shown in (361). The annotation $(\uparrow_i \text{QFOC}) = \downarrow$ indicates they must correspond to interrogative focus in i-structure, undoubtedly with the discourse features $[−\text{GIV}, −\text{ABT}, +\text{PROM}]$ and $+[\text{CONT}]$.

(361) *ima* ‘who’/*nanu* ‘what’:

\[
\begin{align*}
&\text{D} \\
&\quad (\downarrow_i \text{WH}) = + \\
&\quad (\uparrow_i \text{QFOC}) = \downarrow \\
&\quad (\downarrow_i \text{GIV}) = c – \\
&\quad (\downarrow_i \text{ABT}) = c – \\
&\quad (\downarrow_i \text{PROM}) = c + \\
&\quad (\downarrow_i \text{CONT}) = c + 
\end{align*}
\]

Because of these inherent lexical properties, interrogative words tend to fit into the syntactic position with identical properties, like the nominal predicate of pseudo-cleft sentences in Squiliq Atayal. It explains why all of the interrogative sentences in our naturalistic data have pseudo-clefted interrogative words. In-situ interrogative words are so infrequent that they only occur in the grammaticality-judgment tests owing to the fact that although syntactic positions like VP-complement (i.e. OBJ in f-structure) or IP-adjunct (i.e. OBL in f-structure) are compatible with the lexical properties of interrogative words, they are not as favorable as the nominal predicate of pseudo-cleft sentences. Both VP-complement and IP-adjunct correspond to $[±\text{GIV}, ±\text{ABT}, −\text{PROM}]$ in i-structure, the negative value of PROM in which is not obligatorily constrained, so the lexical properties of interrogative words can override it. In order to substantiate the discourse feature $[+\text{PROM}]$ in an unmarked position, prosodic prominence is inevitable. Cross-linguistically in-situ interrogative words all carry prosodic prominence (Mycock 2006). Informants always raised the pitch of in-situ interrogative words in the grammaticality-judgment tests.
8.3.3.3 Transitional Adverbials

Although the regular mapping between c-structure, f-structure and i-structure makes pseudo-clefting a construction specialized for contrastive prominent information, as displayed in Section 8.2.2.1, the usage of pseudo-clefting is extended to accommodate transitional adverbials, which are no longer contrastive information, as exemplified in (316) and (317), repeated below:

(316) nanu’ yasa qu’ syuk-un=nha’ zipun.
Par thus Nom retaliate-PV=3PObl Japanese
‘Therefore, Japanese were retaliated against by them.’ <N>

(317) yasa qu’, m-zyup squ’ ngbang qani. (FLA)
thus Nom AV-fall.into Loc cave this
‘As a result, you fell into the cave.’

The hypothetical diachronic development is formalized in (362).

(362) a. syntactically-assigned properties:
S \rightarrow DP \rightarrow DP

(↑ PREDLINK) = ↓
(↓, PROM) = c +
(↓, CONT) = c +

b. lexically-marked properties:
S \rightarrow DP \rightarrow S

(↑ ADJ) = ↓
(↓, PROM) = c +
(↓, CONT) = −

c. lexically-marked properties:
Particle
(↑, PROM) = c +
(↑, CONT) = −
The first stage is what has been shown for pseudo-clefted contrastive information. At this stage, contrast and prominence are granted to a word through its entering the regular mapping between c-structure and i-structure of pseudo-clefting sentences. Put differently, these two discourse features are not lexically specified but syntactically granted. For some predicates, such as yasa ‘thus/in this way’, they entered this regular mapping just like other predicates in the beginning, mainly to show their prominence and contrast with other ways.

However, the meaning of these predicates is closely related with preceding sentences in terms of the cause-effect relation or the temporal relation so this function of discourse cohesion was gradually highlighted and the predication relationship as well as contrast gradually faded out through repetition. Eventually, the second stage emerged. At this stage, adjunction and modification replaced predication in both c-structure and f-structure, and contrast was no longer granted through their regular mapping with i-structure. The cases still with the nominative case marker qu’ has achieved this stage.

What is more, the omission of case marking keeps pushing forward these transitional adverbials on the way of grammaticalization to discourse particles. For the cases without the nominative case marker qu’, their relationship with the sentential subject is not marked any more. It is believed that they will gradually lose the grammatical function of modification and the discourse feature prominence will be specified in their lexical entries since they have existed in the regular mapping of pseudo-clefting for such a long time that it is impossible for native speakers to differentiate whether prominence comes from pseudo-clefting or from the words themselves. This is the third stage.

It is hypothesized that transitional adverbials which seem to be accommodated in pseudo-clefting underwent the above diachronic development. They are not contrastive any more but are still prominent in drawing the addressee’s attention to following
sentences. Through their frequent occurrence in the regular inter-modular mapping, the discourse feature prominence has become part of their lexical properties.

8.4 Mapping from I-Structure

Through the discussion in Chapter 6, Chapter 7 and the previous sections of this chapter, correspondences in unmarked clauses and marked constructions of GTOP and GFOC have been investigated. Most of discourse functions have been covered, including contrastive topic (CSTOP), switch topic (SWTOP), contrastive theme (CSTHM), scene-setting topic (SSTOP), continuing topic (CTTOP), background information (BKINF), contrastive focus (CSFOC), interrogative focus (QFOC), presentational focus (PSFOC) and completive information (CPINF). Their correspondences to f-structure and to c-structure are recapitulated in Table 8-2. Those derived from the conversion function is marked with (Q). This section will investigate three further issues on the mapping from i-structure: first, the two unmentioned discourse functions so far in the present thesis, ITFOC and HLRHM; second, the distinction between primary CTTOP and secondary CTTOP; last, the mapping theory from i-structure to f-structure.
Table 8-2 Correspondences from Discourse Functions

<table>
<thead>
<tr>
<th>i-structure</th>
<th>f-structure</th>
<th>c-structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrastive Topic (CSTOP)</td>
<td>GOF2</td>
<td>DP</td>
</tr>
<tr>
<td>Switch Topic (SWTOP)</td>
<td>GOF1</td>
<td>CP-ADJ</td>
</tr>
<tr>
<td>Contrastive Theme (CSTHM)</td>
<td>GOF2</td>
<td>DP</td>
</tr>
<tr>
<td>Scene-setting Topic (SSTOP)</td>
<td>GOF1</td>
<td>CP-ADJ</td>
</tr>
<tr>
<td>Continuing Topic (CTTOP)</td>
<td>SUBJ</td>
<td>IP-SPEC</td>
</tr>
<tr>
<td></td>
<td>OBJ</td>
<td>VP-COMP</td>
</tr>
<tr>
<td></td>
<td>OBL</td>
<td>IP-ADJ</td>
</tr>
<tr>
<td>Background Information (BKNF)</td>
<td>SUBJ (Ϙ)</td>
<td>IP-SPEC (Ϙ)</td>
</tr>
<tr>
<td></td>
<td>OBJ</td>
<td>VP-COMP</td>
</tr>
<tr>
<td></td>
<td>OBL</td>
<td>IP-ADJ</td>
</tr>
<tr>
<td>Contrastive Focus (CSFOC)</td>
<td>GOF2</td>
<td>DP</td>
</tr>
<tr>
<td>Introductory Focus (ITFOC)</td>
<td>SUBJ</td>
<td>IP-SPEC</td>
</tr>
<tr>
<td>Interrogative Focus (QFOC)</td>
<td>GOF2</td>
<td>DP</td>
</tr>
<tr>
<td>Highlighted Rheme (HLRHM)</td>
<td>SUBJ</td>
<td>IP-SPEC</td>
</tr>
<tr>
<td>Presentational Focus (PSFOC)</td>
<td>SUBJ (Ϙ)</td>
<td>IP-SPEC (Ϙ)</td>
</tr>
<tr>
<td>Complete Information (CPINF)</td>
<td>SUBJ (Ϙ)</td>
<td>IP-SPEC (Ϙ)</td>
</tr>
<tr>
<td></td>
<td>OBJ</td>
<td>VP-COMP</td>
</tr>
<tr>
<td></td>
<td>OBL</td>
<td>IP-ADJ</td>
</tr>
</tbody>
</table>

8.4.1 ITFOC & HLRHM

Introductory focus (ITFOC) and highlighted rheme (HLRHM) both convey non-contrastive new prominent information, and they differ in that the former is important to subsequent discourse while the latter is unimportant. In Squliq Atayal, these two discourse functions are presented through the marked existential construction.
which is formed by existential verbs always translated as ‘exist’ or ‘there is/are’, like maki’, ungate, cyux, and nyux. Among these existential verbs, only ungate conveys the negative meaning ‘not exist’. They are illustrated in (179) and (250), examples repeated from earlier discussion.

(179) m-aki’ qutux kinhgan rarl mga…
   AV-exist one elder.person past Par
   ‘There was an old woman in the past…’ <N>

(250) tanux=nya’ ga, ngasal balay, ungate balay kyalan.
   front.yard=3SGen Top house real Neg really word
   ‘(Because of/Speaking of) its front yard, a good house, (there are) really no comments.’ <N>

The former is the first sentence in a story, serving to introduce the main character qutux kinhgan ‘one elder person’ into the discourse. This new information, which is conveyed by the grammatical subject, is non-contrastive but important, and the existence sentence enhances its prominence. Hence, it corresponds to ITFOC in i-structure. The latter is a sentence excerpted from the initial part of a narrative. It contains the negative existential verb ungate as the main predicate. The grammatical subject kyalan ‘word’ appears in the discourse for the first time but also for the last time. Its information is non-contrastive prominent but unimportant to subsequent discourse.

The marked constructions of these two existential sentences are respectively realized as the phrase structure rules in (363) and (364).

125 Although the existential verb maki’ is prefixed with the AV marker, the other voice markers do not serve to introduce new important information into discourse.
The phrase structure rule in (363) specifies that the predicate footnode must be *maki’, cyux, or nyux, as indicated in \((\downarrow \text{PREDFN}) = \text{‘maki’/cyux/nyux’}\). In addition, it also stipulates that the IP-specifier must be the grammatical subject in f-structure and the introductory focus in i-structure, as shown in \((\uparrow \text{SUBJ}) = \downarrow\) and \((\uparrow_i \text{ITFOC}) = \downarrow\). The second phrase structure rule in (364) specifies *ungat as the predicate footnode of the whole construction, as indicated in the annotation \((\downarrow \text{PREDFN}) = \text{‘ungat’}\); moreover, the IP-specifier must correspond to the highlighted rheme in i-structure, as marked in \((\uparrow_i \text{HLRHM}) = \downarrow\).

### 8.4.2 Primary vs. Secondary CTTOP

In Table 8-2, CTTOP is the discourse function with the most correspondences to f-/c-structure, excluding those with the application of \(Q\). It corresponds to the most essential three grammatical functions in unmarked clauses, SUBJ, OBJ and OBL. The

---

126 A predicate footnode is the grammatical function of a predicate without arguments in f-structure, corresponding to the head of a phrase in c-structure (Kaplan & Maxwell 1996, King 1997). Technically only the form of the head of IP is checked so PREDFN is adopted here in this equation. The conventional abbreviation of a predicate footnode is PREDFN. In the present thesis, however, its blank space is removed to be in accordance with PREDLINK.
prevalence of CTTOP is not without any reason, and it is due to its own attributes that among the discourse functions with nonprominent information in high-frequency unmarked clauses, CTTOP not only connects to previous discourse but also lays the groundwork for subsequent one; in other words, CTTOP is the most crucial device for discourse coherence. It is proposed in this section that CTTOP is divided into two subtypes: the CTTOP corresponding to SUBJ differs from the CTTOP corresponding to OBJ or OBL in Squiliq Atayal. The former is called “primary continuing topic” (CTTOPπ) while the latter “secondary continuing topic” (CTTOPθ). CTTOPπ has high priority over CTTOPθ in discourse coherence. The evidence comes from discourse-pragmatics itself.

The following sentences in (365) and (366) are excerpted from a family history, where the whole discourse topic generally surrounds the informant’s family.

(365) m-nayang ru’ m-uya ruma’.  
AV-remove and AV-plant bamboo  
‘(We) removed (other plants) and planted bamboos.’ <N>

(366) ruma’ hiya’ ga, sswe’ m-ha, kbalay=ta’, ake, bamboo Emp Top sibling AV-say AV-make=1PObl Par s-kbalay=ta’ ngasal Ø.  
CV-make=1PObl house Ø  
‘(Speaking of) bamboos, (my) siblings said, “Let’s build, Let’s build a house with them.’’” <N>

Thus, as shown in the first sentence, the missing subjects are “we”, referring to the informant and her family. This sentence introduces the new item ruma’ ‘bamboo’ into discourse-pragmatics for the first time. Because it is an object, which does not correspond to PSFOC in i-structure, it has to be topicalized when it is foregrounded, as shown in the beginning of the second sentence. After this GTOP, the next occurrence of ruma’ ‘bamboo’ will become a CTTOPπ, which is realized as the subject of the CV predicate skbalay ‘make/build’, as marked with Ø at the end. Meanwhile, the actor -ta’
‘we’ becomes a NAV oblique. Note that the second sentence does not involve the subject-only constraint of topicalization since the GTOP \textit{ruma} ‘bamboo’ and the subject \textit{Ø} are intervened with one clause \textit{sswe’ mha} ‘(my) siblings said’ and one slip of tongue \textit{kbdelay-ta} ‘we make’.

Once again the above evidence supports the fact that the multiple voice system of Squliq Atayal belongs to the information-salience voice system (Cf. Chapter 6). The existence of LV and CV aims to promote locations and circumstances (including instruments and beneficiaries) to the position of CTTOP in i-structure. Although the subject-only constraint is undoubtedly an instance of the Relational Hierarchy in Keenan & Comrie (1977), where SUBJ is the topmost core argument for syntactic operations, SUBJ and CTTOP are two sides of the same coin in Squliq Atayal (Cf. Chapter 6). From the perspective of discourse-pragmatics, the subject-only constraint manifests that clause-internal information chaining initiated by SWTOP (i.e. GTOP) must end at CTTOP, not CTTOP, as illustrated in (367).

(367) Clause-Internal Information Chaining

\[
\begin{align*}
\text{SWTOP} & \quad \begin{bmatrix} \text{GIV} & + \\ \text{ABT} & + \\ \text{PROM} & + \\ \text{CONT} & - \end{bmatrix} & \cdots & \text{CTTOP} & \begin{bmatrix} \text{GIV} & + \\ \text{ABT} & + \\ \text{PROM} & - \\ \text{CONT} & - \end{bmatrix} \\
\text{SWTOP} & \quad \begin{bmatrix} \text{GIV} & + \\ \text{ABT} & + \\ \text{PROM} & - \\ \text{CONT} & - \end{bmatrix} & \cdots & \text{CTTOP} & \begin{bmatrix} \text{GIV} & + \\ \text{ABT} & + \\ \text{PROM} & + \\ \text{CONT} & - \end{bmatrix}
\end{align*}
\]

In terms of topicalization, Squliq Atayal displays the harmony between the Relational Hierarchy and the clause-internal information chaining since SUBJ and CTTOP are always conveyed by the same argument. As Chapter 9 will present, some languages like Tsou (a Formosan language) instantiate the conflict between them, which further leads to the split-subjecthood phenomena in some Austronesian languages.
8.4.3 Informational Mapping Theory

One-to-many mapping in our theoretical framework represents the overall pattern of the correspondences between i-structure and f-structure, also displaying the complexity of the interactions between syntax and discourse-pragmatics. However, it does not account for how individual correspondence is established. In addition, as mentioned in Chapter 2, most of LFG models in literature have never touched upon the issue on how i-structure is directly mapped with f-structure, only with reliance on the phrase rules in c-structure as the intermedia between i-structure and f-structure. In fact, this is also what has been shown in the present thesis. Therefore, the present thesis puts forth a linking theory between discourse functions and grammatical functions as a necessary complement.

One-to-many mapping and the linking theory are in the type-token relationship, with the former showing the overall pattern while the latter showing the individual case. In LFG, there is a linking theory between a-structure and f-structure called “Lexical Mapping Theory” (Cf. Chapter 3). In order to make a distinction, we call the linking theory between i-structure and f-structure is called as “Informational Mapping Theory (IMT)”. The validity of IMT proposed here still needs to be examined cross-linguistically since Squliq Atayal is the only empirical basis in the present thesis without any comparative evidence from other languages. Although this is very limited, we suggest it as a point of departure.

Before beginning to develop the mechanisms of IMT, we have to be aware of the essential differences between discourse functions and thematic functions. First, the possible number of discourse functions in a clause is not as much restricted as the number of thematic functions, which is restricted by the valence of a verb/predicate. However, it does not mean that the number of discourse in a clause is completely without any restrictions. Its restrictions usually come from other modules, such as the
language-particular constraint on the number of GOF in a clause, and the number of thematic functions a verb subcategorizes. Second, it often happens that the same type of discourse function occurs more than once in a clause, which is unusual to grammatical functions and thematic functions. For instance, there are many cases of two CTTOPs co-occurring in a clause but it is impossible for a verb to subcategorize two agents. However, as mentioned in the first point, the number of identical discourse functions cannot be unlimited because of the limited number of grammatical functions and thematic functions they correspond to.

Third, most important of all, all co-occurrence patterns of discourse functions for core arguments in unmarked clauses are possible—because of the limitation of naturalistic data, only the unmarked clauses with two core arguments are taken into discussion here. At least, most of them are attested in our naturalistic data. It is not out of expectation that a few of the unattested patterns will be found if the size of our naturalistic data is enlarged. In contrast, some co-occurrence patterns of thematic functions are unusual, for instance, that of location and beneficiary. Even though almost all co-occurrence patterns are available, only 6 out of the 12 possible patterns in AV clauses and 7 out of the 12 possible patterns in NAV clauses have more than five tokens, as presented in Table 8-3 and Table 8-4, respectively. They cover most of the tokens observed in our naturalistic data: 93.3% of the tokens in AV clauses and 96.7% of the tokens in NAV clauses. By looking into them, we are able to clarify how discourse functions are associated with grammatical functions. In these tables, the first two columns are the discourse functions of subject and object/oblique. The third and fourth columns are the number and percentage of each co-occurrence pattern. The last column

---

127 There are only 12 clauses with three core arguments in our naturalistic data. The tokens are so few that it is impossible to make an overall description of their co-occurrence patterns of discourse functions.

128 Based on the studies in Chapter 6, subjects correspond to four kinds of discourse functions, CTTOP, BKnINF, PSFOC and CPINF while objects and obliques correspond to only three kinds of discourse functions, CTTOP, BKnINF and CPINF. Given there are only two core arguments in a clause, AV and NAV clauses have $4 \times 3 = 12$ possible co-occurrence patterns.
is the ranking of these patterns, with the most frequent pattern ranked as 1 and with the least frequent one ranked as 6/7. Notice that the major co-occurrence patterns of AV clauses are not exactly the same as those of NAV clauses, which indicates that their linking ways are not completely identical to each other, either.

Table 8-3 *DF Co-occurrence Patterns of AV Clauses*

<table>
<thead>
<tr>
<th>No.</th>
<th>DF Co-occurrence Patterns</th>
<th>Quantity</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CTTOP CPINF</td>
<td>64</td>
<td>35.8</td>
</tr>
<tr>
<td>2.</td>
<td>CTTOP CTTOP</td>
<td>45</td>
<td>25.1</td>
</tr>
<tr>
<td>3.</td>
<td>CTTOP BKINF</td>
<td>28</td>
<td>15.6</td>
</tr>
<tr>
<td>4.</td>
<td>BKINF BKINF</td>
<td>15</td>
<td>8.4</td>
</tr>
<tr>
<td>5.</td>
<td>BKINF CTTOP</td>
<td>8</td>
<td>4.5</td>
</tr>
<tr>
<td>6.</td>
<td>BKINF CPINF</td>
<td>7</td>
<td>3.9</td>
</tr>
</tbody>
</table>

167/179 93.3

Table 8-4 *DF Co-occurrence Patterns of NAV Clauses*

<table>
<thead>
<tr>
<th>No.</th>
<th>DF Co-occurrence Patterns</th>
<th>Quantity</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CTTOP CTTOP</td>
<td>86</td>
<td>39.3</td>
</tr>
<tr>
<td>2.</td>
<td>CTTOP BKINF</td>
<td>16</td>
<td>7.3</td>
</tr>
<tr>
<td>3.</td>
<td>CTTOP CPINF</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>4.</td>
<td>BKINF CTTOP</td>
<td>49</td>
<td>22.4</td>
</tr>
<tr>
<td>5.</td>
<td>BKINF BKINF</td>
<td>22</td>
<td>10.0</td>
</tr>
<tr>
<td>6.</td>
<td>PSFOC CTTOP</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>7.</td>
<td>CPINF CTTOP</td>
<td>9</td>
<td>4.1</td>
</tr>
</tbody>
</table>

208/219 96.7
The problem IMT is faced with is that every co-occurrence pattern goes. Take the co-occurrence of CTTOP and BKINF for example. Table 8-4 shows either of them can correspond to subject and the other will correspond to oblique: the second pattern is a combination of subject-CTTOP and oblique-BKINF and the fourth pattern is one of subject-BKINF and oblique-CTTOP. Linking alone, a mechanism used in LMT, is not enough to achieve the outcome presented in the above tables since there is no fixed linking only between some discourse function, for example, CTTOP, and a grammatical function. It follows that we have to add another mechanism into IMT—that is, choice-making. The employment of both linking and choice-making makes it possible to account for the major patterns in Table 8-3 and Table 8-4.

The application of IMT in Squiq Atayal contains eight steps, as demonstrated in (368) and Figure 8-2. Among these steps, Step 1, 3 and 5 are steps of choice-making (marked with arrows in Figure 8-2), just like crossroads where people face two or more options simultaneously and each of the options leads to a different route. The other steps specify the processes of linking discourse functions with grammatical functions.
Illustration of Discourse Functions

Step 1: Initial Choice of Discourse Functions
- Prominence Mapping: [+PROM]
- Nonprominence Mapping: [−PROM]

Step 2: Prominence Mapping
- Substep 1: [+CONT] \(\rightarrow\) GOF2
- Substep 2: [+GIV] \(\rightarrow\) GOF1
- Substep 3: [−GIV] \(\rightarrow\) SUBJ

Step 3: Midway Choice of Nonprominence Mapping
- General Mapping
- Alternative Mapping

Step 4: General Mapping
- Substep 1: CTTOP, \(\rightarrow\) SUBJ
- Substep 2: Others \(\rightarrow\) NONSUBJ

Step 5: Midway Choice of Alternative Mapping
- AV
- NAV
- Remnant

Step 6: AV
- Substep 1: [+GIV] \(\rightarrow\) SUBJ
- Substep 2: Others \(\rightarrow\) OBJ

Step 7: NAV
- Substep 1: [+GIV] \(\rightarrow\) OBL
- Substep 2: Others \(\rightarrow\) NONOBL

Step 8: Remnant Mapping
Figure 8-2 Procedure of IMT in Squiliq Atayal

Step 1: Discourse Functions

Step 2: Prominence Mapping
- [+CONT]
- [-CONT]

Step 3: Nonprominence Mapping

Step 4: General Mapping
- CTTOP
- Others
- Nonsubject

Step 5: Alternative Mapping
- AV
- NAV
- Remnant Mapping

Step 6: Step 5
- [+GIV]
- Others
- Nonsubject

Step 7: Step 5
- [+GIV]
- Others
- Oblique

Step 8: Step 5
- [+GIV]
- Others
- Nonoblique
In the very beginning of IMT, the first step is to differentiate prominent discourse functions from nonprominent discourse functions. If a discourse function is with [+PROM], then it is fitting to the application of Step 2; if not, Step 2 will be skipped since the application of IMT goes to another route in nonprominence mapping. Step 2 consists of three substeps. The first one links contrastive discourse functions with GOF2 (i.e. GFOC). If a discourse function is not contrastive, then it will enter the second substep, where it will be linked to GOF1 (i.e. GTOP) if it carries [+GIV]. If a discourse function does not meet the above conditions, carrying [+PROM, − CONT, −GIV], it will go to the third substep to be linked to SUBJ. Prominence mapping belongs to marked clauses, the phrase structure rules of which are usually annotated with correspondence constraints. In terms of theoretical economy, it may be removed from our framework to reduce redundancy. However, the proposal of prominence mapping benefits future research on typological comparisons and the universal IMT. If there is a universal principle of prominence mapping, then it is the annotations of correspondence constraints that will be removed. Whether it remains in or be removed from LFG should be left to further examinations.

Step 3 is a choice-making step determining the application of IMT goes to either the general mapping or the alternative mapping. If general mapping is chosen, Step 4 will be applied. The default mapping between CTTOP₂ and SUBJ is satisfied in Step 4, and other discourse functions will be mapped to the other grammatical function (OBJ/OBL) in the second substep.

As mentioned in previous chapters, discourse-pragmatics differs from morphosyntax in that pre-existing rules can be flexibly broken to meet contextual needs, such as the violation of conversational maxims leading to conversational implicature (Cf. Levinson 1983, Mey 1993, Thomas 1995, Verschueren 1999). Thus, with the flexibility of discourse-pragmatics, syntactic interactions with i-structure cannot be as
regular and strict as pure morphosyntactic rules. The alternative mapping in Step 5 offers this flexibility when Step 4 is not applied and the default mapping between CTTOP and SUBJ is not satisfied.

As a choice-making step, Step 5 consists of three choices for alternative mapping: AV, NAV and remnant mapping. If any one of them is chosen, their relevant step will be applied. For AV clauses, alternative mapping will link discourse functions with [+GIV] to SUBJ and others to NONSUBJ, as shown in Step 6. For NAV clauses, alternative mapping will link discourse functions with [+GIV] to OBL and others to NONOBL, as shown in Step 7. Why does the alternative mapping of NAV clauses start with the linking to oblique? Let’s look back at Table 8-4. The fourth, fifth, sixth and seventh patterns belong to alternative mapping, all with discourse functions of oblique carrying [+GIV]. The most straightforward mechanism is to start with the linking to oblique. In addition, it displays a contrast with the general mapping, also manifesting the competition between SUBJ and OBL. This competition will eventually bring about diachronic change, the details of which will be left to Chapter 9. The remnant mapping covers the co-occurrence patterns with fewer than five tokens and cannot be neatly formalized for the moment.

8.5 Summary

In addition to the introduction to the properties of focus construction, this chapter shows that GFOC is the nominal predicate of pseudo-cleft sentences in c-structure and it corresponds to PREDLINK and GOF2 in f-structure. It is argued that the double-tier analysis of verbless clauses is fitting to pseudo-clefting in Squliq Atayal. What is more

129 Even syntactic rules can be violated when syntax meets discourse-pragmatics. In English, a relative pronoun must agree with its head noun in terms of humanness: who(m) for human and which for non-human. The following example in (i) is ungrammatical in syntax, but it can be perfectly acceptable if it is in a felicitous context where the speaker and his/her cat have an intimate relation (Mey 1993).

(i) My cat, who believes that I’m a fool, enjoys tormenting me. (Mey 1993:26)
complex, GFOC corresponds to four discourse functions in i-structure, CSTOP, CSTMH, CSFOC and QFOC. Among its tokens, those with given information are much more than those with new information so it is proposed that GFOC should be viewed as a syntactic position of contrast rather than a syntactic position of focus (new information) in Squiliq Atayal. Both GTOP and GFOC attest the Prominence-Fist Constraint has the top priority in the ranking of i-structural constraints, over the Aboutness-Last Constraint and Given-First Constraint.

At the end of this chapter, it is argued that the discourse functions ITFOC and HLRHM are attested in existential sentences. In addition, there is a distinction between CTTOP_{x} and CTTOP_{θ} in terms of discourse coherence. Based on the mapping between discourse functions and grammatical functions, we make a proposal of IMT, which employs both of linking and choice-making, as the complement of one-to-many mapping. The mapping between i-structure and f-structure in LFG can be direct without resort to phrase structure rules in c-structure.
9. 

Split-Subjecthood: 
Diachronic Hypothesis

Time does not cease. Change does not end. 
--Hal Botland

9.1 Overview

Split-subjecthood is one of the remarkable characteristics of Austronesian languages. Based on the studies in the previous chapters and the comparisons between Squiliq Atayal and another Formosan language Tsou, this chapter will propose a hypothesis about the diachronic development of split-subjecthood particularly in Austronesian languages. The similarities and differences between these two languages prove that split-subjecthood is associated with the interactions between syntax and information structure, also showing that the force of discourse-pragmatics gradually permeates into syntax over time through the mapping between different modules in LFG, especially f-structure, a-structure and i-structure.

9.2 Split-Subjecthood

What is split-subjecthood? How is it substantiated in linguistic typology? This section will offer the answers to these two questions. The split-subjecthood phenomena in Squiliq Atayal and Tsou will be demonstrated in particular.
9.2.1 Notion

The notion of grammatical subject was challenged by early linguistic studies on Austronesian linguistics, such as McKaughan (1973) and Schachter (1976). They argued that the notion of subject was inapplicable since the nominative argument in Philippine-type languages was a topic per se. For them, subject and topic were two mutually independent notions in syntax. Later linguistic studies on Austronesian languages however discover that the morphosyntactic properties of grammatical subject attested in European languages do exist in Austronesian languages and that unlike European languages, they are not all carried by nominative arguments, or rather some of them are carried by actors when in NAV clauses (Arka 2003, Keenan 1976, Kroeger 1993, Schachter 1976, inter alia). This is so-called “split-subjecthood”. To be more specific, nominative arguments in AV clauses, which convey the thematic function ACTOR, have the full set of morphosyntactic properties of grammatical subject, just as nominative arguments display in European languages; on the other hand, nominative argument and oblique actor in NAV clauses respectively carry some of these properties, with subjecthood split up into two arguments.\(^\text{130}\) In addition to Austronesian languages, split-subjecthood is also attested in syntactically ergative languages, such as Dyirbal, in which both the ergative argument and the absolute argument in transitive clauses have some morphosyntactic properties of subject in traditional grammar (Dixon 1972, 1994, Manning 1996b).

How to account for split-subjecthood has been intriguing to theoretical syntacticians. Syntacticians under the transformational approach resort to the difference

\(^{130}\)In Germanic linguistics, “[t]he term ‘oblique subject’ refers to an NP that behaves syntactically like a subject without having nominative case (Faarlund 2001:99).” Although this term may be applied to the oblique actor in NAV clauses in Austronesian languages since the oblique actor has some morphosyntactic properties of grammatical subject, it may incorrectly imply that the co-existing nominative argument in NAV clauses does not have any morphosyntactic properties of grammatical subject.
in configurational positions. For instance, Guilfoyle, Hung & Travis (1992) claims that split-subjecthood results from the association of VP-specifier with theta-sensitive properties and that of IP-specifier with subject-sensitive properties. Noticing the correlation of split-subjecthood with information structure, M. Chang (2004) further strengthens the argumentation of positional difference with the movement to the left periphery.

Instead of attributing the explanatory power to configurationality, LFG syntacticians establish a new grammatical function independent from SUBJ in f-structure. Kroeger (1993) analyzes NAV oblique in Tagalog as having the grammatical function ACTOR in f-structure. In addition, taking into consideration the phenomena on ergativity discussed in Dixon (1979, 1994), Manning (1996b), Marantz (1984), etc., and those on Philippine-type languages discussed in Schachter (1976), Kroeger (1993), etc., at the same time, Falk (2006) divides the grammatical function SUBJ into two, GF and PIV, since subject properties can be grouped into two types, which are listed in (369) and (370) respectively. Subject properties of the first type are carried by the actor of intransitive clauses (S) and that of transitive clauses (A) whereas subject properties of the second type are carried by the actor of intransitive clauses (S), and the patient of transitive clauses (P) in syntactically ergative languages or the nominative argument in Philippine-type languages.\(^{131}\) Syntactically ergative languages and Philippine-type languages are labelled together as mixed-subject languages.

(369) Type 1 subject properties: GF (Falk 2006:16)
- S/A
- Agent agreement in the active voice
- Most likely covert/empty argument
- The addressee of imperatives
- Anaphoric prominence

\(^{131}\) The symbols S, A and P used in Falk (2006) come from Dixon’s (1994) convention.
Switch-reference systems
Controlled argument (PRO) (in some languages)
Discourse topic

(370) Type 2 subject properties: PIV (Falk 2006:16)
- S/P in syntactically ergative languages; Nominative in Philippine-type languages
- Shared argument in coordinated clauses
- Controlled argument (PRO) (in some languages)
- Raising
- Extraction properties
- Obligatory element
- “External” structural position
- Definiteness or wide scope

Falk’s distinction between $\overline{GF}$ and PIV suggests that AV clauses in Philippine-type languages behave like active clauses in European languages since their subject conveys both PIV and $\overline{GF}$, while NAV clauses in Philippine-type languages behave like ergative clauses in ergative languages since their subject conveys $\overline{GF}$ and their oblique conveys PIV. This accords with the claim in H. Chang (1997) that two Formosan languages, Seediq and Kavalan, are split-ergative languages, with AV clauses behaving like active clauses while NAV clauses behaving like ergative clauses.

Whether split-subjecthood theoretically results from the division of morphosyntactic subject properties into two different configurational positions or two different grammatical functions, empirical evidence shows that it is not a single homogeneous phenomenon with one identical division of subject properties in every language but a collection of heterogeneous phenomena with the division of subject properties differing from language to language. This is manifested in the next section with the evidence from Squilq Atayal and Tsou.
9.2.2 In Tsou & Squliq Atayal

By comparing Tsou and Squliq Atayal, two Formosan languages, this section will demonstrate that even within languages with split-subjecthood the scale of split-subjecthood varies from language to language.

9.2.2.1 In Tsou

All of the data and the linguistic phenomena in Tsou presented in the present thesis are cited or slightly modified from previous linguistic studies. Although Tsou is a Formosan language, it is like a Philippine-type language, which is characterized by split-subjecthood. Like Squliq Atayal, Tsou has a rigid unmarked word order Verb-Object/Oblique-Subject (H. Chang & Pan 2016, Zeitoun 2000). Here the subject refers to the argument marked with the nominative case. It is illustrated in the examples (371) and (372), where the sentence-final arguments follow the same (nominative) case marker ‘o.

(371) m-o b-aïto to oko 'o ino. (M. Chang 2004:36)\textsuperscript{132}
AV-Rea AV-see Obl child Nom mother
‘The mother watches the child.’

(372) Ø-i-si ait-i to ino ‘o oko. (M. Chang 2004:37)
NAV-Rea-3S see-LV Obl mother Nom child
‘The child is seen by the mother.’

In the former, both the auxiliary mo and the verb bai’to ‘see’ carry AV, which indicates the subject is an actor, and there are two arguments following the predicate, oko ‘child’

\textsuperscript{132} In M. Chang’s thesis on Tsou, case markers are termed as “locational determiners”. In order to be in accordance with other studies on Tsou and the examples from Squliq Atayal, this term is replaced with the glosses of case markers.
and *ino* ‘mother’. The preceding argument *oko* ‘child’ is the object while the sentence-final argument *ino* ‘mother’ is the subject since the English translation tells us that the sentence-final argument *ino* ‘mother’ is the actor of this event, meeting the requirement of the voice marking on the predicate. In contrast, the latter example has a NAV-marked auxiliary *isi* and a LV-marked verb *aiti* ‘see’, which requires that the subject must be the non-actor of the event. Its English translation shows that the sentence-final argument *oko* ‘child’ is the non-actor, so it is the grammatical subject with the theta-agreement with the voice marking. Meanwhile, the actor *ino* ‘mother’ is placed in the sentence-medial position.

Nominative arguments in Tsou are the only argument which can undergo relativization (M. Chang 1998a, 2004). According to Keenan & Comrie (1977), relativization must comply with the relational hierarchy, in which grammatical subject is the candidate with the topmost priority to be relativized. In other words, if a language only permits one argument to be relativized, this argument must be grammatical subject. It is illustrated in the following examples from (373) to (376) that the sentence-final argument in Tsou is a grammatical subject without any doubt.

(373) \[ \text{RC m-o b-aito to oko Ø ci ino} \] (M. Chang 2004:37)

\[
\begin{array}{ll}
\text{AV-Rea} & \text{AV-see Obl child Ø Rel mother} \\
\end{array}
\]

‘the mother who watches the child’

---

133 Unmarked sentences in Tsou must contain one auxiliary, which agrees with the main verb in the theta-agreement with the grammatical subject (Zeitoun 2000). Early works on Tsou like Tung (1964) terms these obligatory auxiliaries as “beginners” or “particles”.

134 Two hierarchies are relevant to this chapter, the relational hierarchy and the thematic hierarchy. The former is applied in f-structure while the latter is applied in a-structure. They are displayed below:

(i) Relational Hierarchy (Keenan & Comrie 1977)
Subject > Direct Object > Indirect Object > Major Oblique > Genitive > Object of Comparison
Agent > Patient/Beneficiary > Instrument > Theme > Path/Location/Reference Object
The examples (373) and (374) differ in their voice marking: the former contains a relative clause marked with AV while the latter contains a relative clause marked with NAV. The nominative arguments, ino ‘mother’ in (373) and oko ‘child’ in (374), are relativized in these two examples, and they are grammatical. However, if the sentence-medial argument, i.e. the non-nominative argument, is relativized, like oko ‘child’ in (375) and ino ‘mother’ in (376), the relative clause will be ungrammatical. The contrast can be demonstrated in the comparisons between (373) and (375), and between (374) and (376). The above evidence shows that nominative arguments, whether in AV clauses or in NAV clauses, are grammatical subjects. Following the relational hierarchy, our LFG model will treat the nominative arguments as SUBJ in f-structure (or PIV in Falk’s (2006) terminology).

After identifying the nominative argument in Tsou as the real grammatical subject in LFG, we are going to look into the phenomena of split-subjecthood, which exist in NAV clauses since the nominative argument is not an actor and the actor competes with the nominative argument for the morphosyntactic properties of subject in traditional grammar. The most typical phenomena of split-subjecthood are the addressee of imperatives and anaphoric prominence, as exemplified below:
Example (377) is an imperative sentence in NAV, but the missing addressee is the actor, i.e. an invisible you, instead of the nominative argument naveu ‘rice’. The other example in (378) contains reflexive binding. In this sentence, the reflexive pronoun yachi ‘self’ is the nominative argument and it is coreferential with the actor pasuya. That is to say, the actor is gifted with the anaphoric prominence in NAV clauses.

Another phenomenon of split-subjecthood comes from control constructions. As demonstrated in M. Chang & Tsai (2001), Tsou runs afoul of the widely-accepted tendency that control complies with the relational hierarchy since the controller is always the actor in both AV and NAV clauses.\footnote{According to Yeh (1997) and Zeitoun (2000), some control constructions in Tsou, especially those with manipulative verbs and perception verbs, have the linker/conjunction ho ‘and’. They will not be discussed here since it is controversial whether they are real control constructions or just coordinations.} In other words, the controller of control constructions in Tsou must be an actor, i.e. either the nominative argument of AV clauses or the actor of NAV clauses. This is exemplified in (379). Notice that the nonfinite verb pabon ‘eat’ is prefixed with the causative morpheme pa-, indicating that its invisible argument (i.e. the controllee) is a causer. The main clause in (379) is a NAV one, and the controller which the invisible controllee is coreferential with is the NAV actor -’o ‘by me/my’, not the grammatical subject taini ‘he’. This is different from what is always seen in European languages. As shown in the English example (380), the controller of to eat must be the grammatical subject the child, not the oblique agent his
(379) i-`o ahiy-a pa-bon-i na taini. (M. Chang & Tsai 2001:9)  
NAV-1S.Gen force-PV Cau-eat-AV Nom 3S.Nom  
‘I forced him to eat.’

(380) **The child** was forced by his mother to eat.

Apparently, the selection of the controllee does not demonstrate split-subjecthood properties, but the selection of the controller does. However, recall that among the properties of $\mathcal{GF}$ in (369), Falk (2006) proposes an optional property that the controlled element (i.e. the controllee) in some languages is an actor. Undoubtedly, Tsou is not one of them.

What is more intriguing, one of the $\mathcal{GF}$ properties in Falk (2006) is about agent agreement in the active voice. The split-subjecthood property is manifested in NAV clauses in Tsou. The verb agreements in Tsou, person, number and visibility, accord with the nominative argument in AV clauses and the actor in NAV clauses, but not with the nominative argument in NAV clauses since it is not an actor, as demonstrated in (381) and (382):

(381) m-i-hin’i/*-ta b-aito to mo’o ’e o’oko. (M. Chang 2004:38)  
AV-Rea-3P/-3S AV-see Obl Mo’o Nom children  
‘The children saw Mo’o.’

(382) Ø-i-hin’i/*-ta ait-i ta o’oko ’o mo’o. (M. Chang 2004:38)  
NAV-Rea-3P/-3S see-LV Obl children Nom Mo’o  
‘Mo’o was seen by the children.’

Both of the above examples contain two arguments, *mo’o* and *o’oko* ‘children’. The
former is singular and the latter is plural in number. The example (381) is an AV clause, and the main verb can only be suffixed with the plural marker -hin'i, but not the singular marker -ta. It is apparent that the verb agrees with the plural argument o'oko ‘children’, which serves as both the actor and the nominative argument of this AV clause. Conversely, the example (382) is a NAV clause, and likewise the main verb can only be suffixed with the plural marker -hin'i. However, at this time the verb agrees with the actor o'oko ‘children’, not with the nominative argument (non-actor) mo'o. Before looking into other phenomena, we have to add that these verb agreement markers undoubtedly originate from clitic pronouns. In some works on Tsou, they are still analyzed as clitic pronouns (Cf. Zeitoun 1992, 2000).

Although relativization and topicalization are viewed as the same mechanism in some previous linguistic works, they differ in that the former proves nominative arguments to be grammatical subjects whereas the latter manifests split-subjecthood in Tsou. Topicalization is sensitive to both nominative arguments and actors (H. Chang & Pan 2016, M. Chang 1997, 1998a, b, Law 2007, Zeitoun 2000). To be more specific, when these two roles are overlapped on the same argument, that is, the nominative argument of AV clauses, it is the only candidate for topicalization. In contrast, when these two roles are carried by two distinct arguments, that is, the nominative argument and the actor of NAV clauses, both of them are candidates for topicalization. This is illustrated in the following examples:

(383) 'e/si/ta pasuya, m-i-ta e-mafe to fou Ø,
Top Pasuya AV-Rea-3S AV-like.to.eat Obl pork Ø
‘Pasuya, he likes to eat pork.’ (M. Chang 1998a:123)
(384) *e/si ta fou, m-i-ta e-mafe Ø, 'e/si ta pasuya.
   Top  pork AV-Rea-3S AV-like.to.eat Ø Nom Pasuya
   Intended for ‘As for the pork, Pasuya likes to eat (it).’ (M. Chang 1998a:123)

(385) 'e/si ta fou, i-ta mafe-a ta pasuya Ø.
   Top  pork NAV.Rea-3S like.to.eat-PV Obl Pasuya Ø
   ‘As for the pork, Pasuya likes to eat (it).’ (M. Chang 1998a:123)

(386) 'e/si ta pasuya, i-ta mafe-a Ø, 'o fuzu.
   Top  Pasuya NAV.Rea-3S like.to.eat-PV Ø Nom pork
   ‘Pasuya, he likes to eat pork.’ (M. Chang 1998a:123)

The examples (383) and (384) are AV clauses, and the examples (385) and (386) are
NAV clauses. In (383), the nominative argument is topicalized, or we may state that the
null nominative argument Ø is coreferential with the GTOP pasuya, so it is grammatical.
Nevertheless, in (384), it is the object Ø that is coreferential with the GTOP pasuya,
which makes this sentence ungrammatical. On the other hand, (385) and (386) do not
have such difference since both of them are grammatical even though the example (385)
is a case of topicalizing nominative arguments while the example (386) is a case of
topicalizing actors. Topicalization is sensitive to both nominative argument and actor in
NAV clauses, unlike other split-subjecthood phenomena.

9.2.2.2 In Squliq Atayal

With regard to split-subjecthood, Squliq Atayal is on the borderline the Philippine
type or may not be viewed as one of this type since split-subjecthood is very limited in
this language. It is only manifested in two phenomena, the addressee of the imperative
and anaphoric prominence.

The following example in (387) illustrates the addressee of the imperative. In this
example, the invisible argument is the actor you, and the nominative argument zyaw
qani ‘this thing’, with a preceding nominative case marker qu’, is still present. This is exactly like what we have seen in Tsou, except for the difference of morphosyntactic marking.

(387) syuk-i cikay qu’ zyaw qani. (s.m. L. Huang 1993:112)
   answer-CV some Nom thing this
   ‘Please answer this thing.’
   (Lit. ‘This thing is answered a little!’)

The other split-subjecthood phenomenon is anaphoric prominence, which is not touched in any previous formal publication on Squiliq Atayal, partly because of the rarity of informants sharp in syntactic/semantic intuition, and partly because of the inference of discourse context. The reflexive pronoun in this language is composed of two elements, a pronoun and the posterior quantifier nanak ‘only/self’. As illustrated in (388), the reflexive pronoun hiya’nanak ‘himself’ is the nominative argument as well as the grammatical subject since it is preceded by the nominative case marker qu’. This reflexive pronoun is coreferential with the actor tali’, a phenomenon in conflict with the relational hierarchy.

(388) bhy-an na’ tali’, qu’ hiya’ nanak.
   hit-LV Obl Tali Nom 3SNom self
   ‘Tali hit himself.’ (Lit. ‘Himself was hit by Tali.’) <E>

If we reverse the positions of the reflexive pronoun and its binder, as exemplified in (389), where the reflexive pronoun -nya’nanak ‘himself’ is the oblique actor and its binder is the nominative argument tali’, the sentence will become ungrammatical. Even topicalizing the binder tali’, which causes the binder to precede the reflexive pronoun -nya’nanak ‘himself’, cannot save the sentence from ungrammaticality, as illustrated in
Based on the above evidence, we can tell that NAV actors have the anaphoric prominence in Squliq Atayal, just like what is shown in Tsou. Reflexive binding in these two languages strictly complies with the thematic hierarchy, in which agent/actor is the candidate with the topmost priority to serve as the binder.

Except for the above two phenomena of split-subjecthood, the other morphosyntactic properties displaying split-subjecthood are all sensitive to nominative argument in NAV clauses (i.e. grammatical subject of NAV clauses) in Squliq Atayal. For instance, according to L. Huang (1993) and L. Huang & Hayung (2016), the grammatical subject in serial verb constructions and pivotal constructions serves as both the controller and the (invisible) controllee at the same time. This can be exemplified in (391), where the nominative argument sayun in the sentence-final position is the controller even though its thematic role is patient and the invisible controllee of the non-finite verb *mihiy ‘hit’ serves as the nominative argument since the AV prefix *m- indicates that the nominative argument/subject is the actor *sayun. Even though *tali’ is the actor of the finite verb *qrgan ‘stop’, it plays no role in the controlling relationship between the finite clause and the embedded nonfinite clause.

\[^{136}\text{In fact, L. Huang (1993) and L. Huang & Hayung (2016) have never made the statement on control because of the framework they adopt. They only propose the AF-constraint, which states that the second verb and the sequent verb can only carry the marker of Actor Focus/Voice no matter what focus/voice marker the first verb carries in serial verb constructions and pivotal constructions. However, their abundant examples, systematic presentations and detailed descriptions have clearly manifested the statement presented in the present thesis.}\]
(391) qrq-an  tali’ ‘m-ihiy kuzing] sayun. (s.m. L. Huang 1993:93)
             stop-LV Tali AV-hit 1SNeu Sayun
            ‘Tali stopped Sayun from hitting me.’

In addition, as elaborated in Chapter 7, topicalization in this language strictly obeys the subject-only constraint so split-subjecthood does not exist either. Moreover, Squliq Atayal resembles Tsou in that only nominative argument can be relativized, a fact already presented in Liu (2004a, b, 2005). It is exemplified below:

(392) siy-on  na’ tali’ qu’ ngasal [ka’ q-niq-an sehuy na’ sayun].
            like-PV Obl Tali Nom house Rel eat-LV taro Obl Sayun
            ‘The house where Sayun ate taros is like by Tali.’ (s.m. Liu 2004b:44)

In terms of thematic role, the head noun ngasal ‘house’ serves as the location in the relative clause. The sentence is grammatical only if the head noun functions as the grammatical subject in the relative clause, as shown in the main verb carrying the LV marking C-...-an (i.e. qniqan).

9.2.2.3 Comparisons

The similarities and differences between Squliq Atayal and Tsou are summarized in the following table:
<table>
<thead>
<tr>
<th></th>
<th>Squiliq Atayal</th>
<th>Tsou</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAV</td>
<td>NAV</td>
<td></td>
</tr>
<tr>
<td><strong>Word order</strong></td>
<td>VOS</td>
<td>VOS</td>
</tr>
<tr>
<td>Nominative case</td>
<td>subject</td>
<td>subject</td>
</tr>
<tr>
<td>Relativization</td>
<td>subject</td>
<td>subject</td>
</tr>
<tr>
<td>Topicalization</td>
<td>subject</td>
<td>subject</td>
</tr>
<tr>
<td></td>
<td>actor</td>
<td></td>
</tr>
<tr>
<td>Verb agreement</td>
<td>subject</td>
<td>actor</td>
</tr>
<tr>
<td>Controller</td>
<td>subject</td>
<td>actor</td>
</tr>
<tr>
<td>Anaphoric prominence</td>
<td>actor</td>
<td>actor</td>
</tr>
<tr>
<td>Addressee of imperatives</td>
<td>actor</td>
<td>actor</td>
</tr>
</tbody>
</table>

The same word order in Squiliq Atayal and Tsou removes the possibility of its influence on split-subjecthood. The above table clearly demonstrates that split-subjecthood varies between Squiliq Atayal and Tsou to such an extent that it is not a homogeneous phenomenon but a collection of various phenomena. For example, although split-subjecthood is manifested in the addressee of imperatives and anaphoric prominence in both of the languages, it is not manifested in topicalization, verb agreement and controller of Squiliq Atayal. Apparently, Tsou has a larger scale of split-subjecthood than Squiliq Atayal. Even so, both case marking and relativization in Tsou still favor the nominative argument, not the actor, in NAV clauses.

Since the scale of split-subjecthood varies from language to language, exemplified in the differences between Squiliq Atayal and Tsou, the syntactic mechanisms of synchronic linguistic research, such as distinguishing the structural position of theta-sensitive properties from that of subject-sensitive properties in MGG or dividing
the original grammatical function SUBJ into two new grammatical functions $\text{GF}$ and $\text{PIV}$ in LFG, do not suffice to make a thorough account of split-subjecthood. To be more specific, if it is the case, as proposed in the synchronic analyses that the actor in NAV clauses satisfies its EPP feature in a certain structural position or carries a special grammatical function, it is difficult to explain why that particular position or grammatical function has greater effects on Tsou than on Squilq Atayal.

9.3 Diachronic Hypothesis

This section presents the diachronic hypothesis about the emergence of split-subjecthood. The development results from the interplay between information structure and syntax. With both the quantitative and qualitative evidence from Tsou and Squilq Atayal, we will verify this hypothesis and its premise.

9.3.1 Syntactic Change

So far, it has been shown that languages vary in the extent of split-subjecthood, which is a collection of morphosyntactic phenomena. The mechanism of dividing syntactic units into two types in previous synchronic linguistic analyses does not work well in its explanatory power. In order to gain a better explanation, it is necessary to take the diachronic perspective into consideration.

However, traditional reconstruction in historical linguistics may not meet our needs because “[m]uch of what is called grammatical reconstruction in the literature is just the plain vanilla comparative method applied to morphemes in the usual way. (Harrison 2003:228)” Although the reconstruction of the voice markers in PAN by comparing them in decedant languages seems to show that PAN was an ergative language or a mixed ergative language, as claimed in Starosta, Pawley & Reid (1982, 2009[1981]), it still cannot make an account of the fact that the decedants of PAN, like Squilq Atayal
and Tsou, have different scales of split-subjecthood no matter how closely ergativity is associated with split-subjecthood.

In addition, it is impossible to attribute all of the syntactic properties observed in decedant languages to their shared ancestor; that is to say, some of the syntactic properties may emerge and develop independently but in the same direction of language change. Take cleft constructions in Germanic languages for example. In Present-Day English, colloquial German and Danish, the clefted constituents are not restricted to DPs only (Bertollo 2014, Luo 2009, Nølke 1984), as exemplified in the following examples from these three languages respectively:

(393) It was **for $500** that Ann bought a first edition of *Naked Lunch.*

(McCawley 1988:59)

(394) Es ist **mit Hans**, dass ich gespielt habe. (Bertollo 2014:226)

It is with Hans that I played

‘It’s with Hans that I played.’

(395) Det var **i Frankrig**, Hansens var sidste sommer. (Nølke 1984:85)

It was in France Hansens was last summer

‘It was in France (that) the family Hansen was last summer.

It is shown that the clefted constituents in the above modern Germanic languages can be PPs, like **for $500** in the English example, **mit Hans** ‘with Hans’ in the colloquial German example and **i Frankrig** ‘in France’ in the Danish example. If there were no historical documents of these languages, the outcome of the syntactic reconstruction with the comparative method would be highly likely to support the existence of cleft

---

137 According to Avery D. Andrews (p. c.), this sentence looks strange to some native speakers of the English language, and its grammaticality may be different from dialect to dialect. However, since clefting PPs is viewed as a grammatical construction in general English by previous studies, such as Baker (1995), McCawley (1988) and Ward, Birner & Huddleston (2002), among others, the present thesis has no choice but to accept this assumption, temporarily sidestepping the issue on dialectal variations.
constructions clefting PPs in the grammar of the ancestor language so that the decedant languages have this syntactic property in common via inheritance. This reconstruction however is not in accordance with the real diachronic development of cleft constructions in the history of English because only DPs could be clefted in Old English, like Petrus in the following example:

(396) Nis hit na Petrus þæt þær cnucað, ac is his ængel. (Mitchell 1985b:102)
not-is it not Peter that there knocks but is his angel
‘It isn’t Peter who is knocking there, but his angel.’

That is to say, the cleft constituents of other categories observed in Present-Day English emerged in the later diachronic development, not being part of the historical inheritance from ancestor language (Ball 1991, Patten 2012); the diachronic development of cleft constructions in Germanic languages independently followed the same direction. Therefore, while applying the comparative method to syntactic reconstruction, linguists should take into consideration the possibility of diachronic development, which could resort to syntactic theories for guidance. Emphasizing the method of internal construction is of the same importance as the comparative method, Lehmann (1979) reminds linguists of the short-sightedness of neglecting any method, or any of the techniques and findings which have led to the advances in descriptive and historical linguistics.

Despite the difference in syntactic theories, the present thesis agrees with the proposal in Faarlund (1990):

In a parallel fashion, language has a double ontology; it is both a mental object and a social object. In its capacity as a mental object it is shaped and conditioned by genetic or biological factors connected with the human mind. As a social object it is shaped and conditioned by its role as a mean of human communication. Thus
linguistic change can and must be explained both on the basis of properties of the human mind and on the basis of properties of human communication and social intercourse. (Faarlund 1990:34)

Language has a dual nature, involving both mental and social domains. “[I]n the social domain we may find answers to why a change took place, and in mental domain we may find answers to why it took the course that it did (Faarlund 1990:45).” However, as for how the social domain is associated with the mental domain, the present thesis rejects the idea that language acquisition is the only channel between these two domains—that is, on the premise that Universal Grammar sets up everything from the beginning, syntactic changes are attributed to the deficient or opaque input from first language acquisition, as proposed by Lightfoot (1979, 1989a, b, 1991, 2003, 2006a, b) and Roberts & Roussou (2003), or result from the competition between two co-existing grammatical systems, as proposed by Kroch (1989), Fontana (1993), Pintzuk (1999, 2002, 2003, 2005) and Pintzuk & Taylor (2006). The former is strongly attacked by historical linguists particularly of the grammaticalization approach since gradual diachronic change has been attested in many cases (Fischer 2007). “[G]radualness poses something of a challenge to the theory of grammar because it characterizes not only shifts in stylistic preferences but also the diffusion of changes in syntactic parameter settings (Kroch 2000:720).” For the latter, the competing grammar hypothesis, although it is able to explain the gradualness of syntactic changes, it is not accepted by the mainstream of MGG syntacticians since this proposal has limited representation on diachronic development and it is against the common belief that there is only one language-particular grammar for each language (Van Kemenade 2007).

The theoretical framework of the present thesis, as elaborated in Allen (2001), has an advantage over any generative theory which treats all change as essentially sudden in that it has the ability to accommodate gradual change as well. “In LFG, a change can be
treated as sudden in the sense that the introduction of a new possibility in the grammar must be sudden but gradual in the sense that it is implemented gradually through lexical entries (Allen 2001:61-62).” What is more important, in additional to lexical properties, the parallelism of modules gives LFG more flexibility in accounting for syntactic change by providing the mental domain and the social domain of language with another channel to connect with each other. To be more specific, syntactic change can be attributed not only to the internal factors of syntax and lexical entries, which are the basics of mental domain of language, but also to the interaction with information structure, which are the basics of human communication in the social domain of language. As stated in Mithun (2003), “the study of an individual syntactic construction in isolation can fail to yield the same kind of understanding that might be possible when the language is considered as a whole. (p. 562)” Givón (1979a) proposes the following diachronic process composed of syntacticization, morphologization, lexicalization, and phonological attrition:

(397) Discourse → Syntax → Morphology → Morphophonemics → Zero

(Givón 1979a:209)

Despite its emphasis upon the diachronic development of linguistic form, the above process also reveals that the influence of information structure, a means of human communication in discourse-pragmatics, permeates into syntax, further into morphology, and even into phonology step by step. Since human communication aims to achieve mutual intelligibility to the greatest extent, the change in information structure must be gradual, further motivating syntactic change gradually through the interaction between syntax and information structure.

The diachronic hypothesis proposed in this chapter cannot be directly supported by historical documentation, which exists only for a few Austronesian languages. We have
to resort to linguistic comparison and reconstruction. Although the accuracy of the comparative and internal reconstruction of syntax always faces challenges from linguists, it is not completely untenable (Fox 1995, Harris & Campbell 1995). What languages are compared is crucial to the validity of our hypothesis. Here we choose Squiliq Atayal and Tsou as the targets of our comparative studies for the following reasons. First of all, these two languages are more closely related to PAN than Extra-Formosan languages in terms of the reconstructed genealogical relationship of the Austronesian family (Blust 1977, 1995, 1999, Ross 2009) (Cf. Chapter 1). Although the preservation of archaic phonological or morphological properties does not guarantee the preservation of syntactic properties, there is no doubt that genetically immediate descendants of PAN, like Tsou and Squiliq Atayal, are more likely to provide us with clues to the diachronic development of split-subjecthood. Second, the availability of data is an important factor to be considered for any study on minority languages. Although Tsou is not the target language of the present thesis, fortunately, there are enough previous studies demonstrating the contour of split-subjecthood and related evidence, as partly shown in Section 9.2.2.1. In the later sections, we will continue making use of the valuable data and facts observed in previous studies. Last, most important of all, our comparison aims to reconstruct the emergence of split-subjecthood, so the two languages under comparison must have some differences pinpointing this issue—that is, the scale of split-subjecthood in Tsou is larger than in Squiliq Atayal (Cf. Section 9.2.2.3). The comparisons between Squiliq Atayal and Tsou offer linguists a clue to the emergence of split-subjecthood in Austronesian languages. In the following sections, the present thesis will propose a diachronic hypothesis to explain why split-subjecthood varies from language to language, by means of the interaction between syntax and information structure.
9.3.2 Development of Split-Subjecthood

Split-subjecthood means that in terms of the morphosyntactic properties of subject in traditional linguistic analyses, some go with the nominative argument (i.e. grammatical subject), some go with the actor, and still others go with both of them. As for how split-subjecthood emerged and developed in history, there are two proposals: one is Cole et al. (1980) and the other is Shibatani (1988b).

Cole et al. (1980) investigates the typological and diachronic evidence from several Germanic languages, Polynesian languages, and (Old and Modern) Georgian, showing that behavioral subject properties are acquired historically prior to subject coding properties. The historical development is divided into three stages. At the early stage, none of morphosyntactic subject properties are associated with oblique arguments. Then, these arguments come to exhibit the syntactic subject properties even though they still carry with non-nominative marking. In fact, this second stage is the stage of split-subjecthood. Finally, these arguments come to possess both syntactic and morphological subject properties; that is, they have the nominative marking. This historical development follows Givón’s (1971, 1979a, b) observation that today’s morphology is yesterday’s syntax. Despite the abundant typological and diachronic evidence in this study, Cole et al. (1980) does not account for why these oblique arguments have the potential for the acquisition of subjecthood.

In an introduction to passive and voice, Shibatani (1988b) notes in passing that in Austronesian languages the oblique actor in NAV clauses gradually lost its subjecthood over time, further leading to split-subjecthood. Instead, it is the nomanitive argument in NAV clauses that acquired subjecthood. That is to say, Shibatani (1988b) and Cole et al. (1980) are contradictory in terms of the diachronic development of split-subjecthood. Although no supporting evidence is presented, the proposal in Shibatani (1988b) cannot
be overlooked since it refers to Austronesian languages in particular.

The diachronic hypothesis proposed by the present thesis accords with Cole et al. (1980), having the opposite view to Shibatani (1988b). That is to say, the nominative arguments in NAV clauses have been real grammatical subjects in history. They are termed “nominative subjects” in the following content, as distinct from oblique actors in NAV clauses in Austronesian languages and oblique subjects in Germanic languages. In contrast, the oblique actors in NAV clauses have never been real grammatical subjects, but they did gradually acquire some morphosyntactic subject properties defined in traditional grammar under the influence of information structure. The oblique actors in NAV clauses in both of Squliq Atayal and Tsou are at the second stage in Cole et al. (1980). However, these two languages further represent two separate stages in the diachronic development of split-subjecthood. The earlier stage is preserved by Squliq Atayal while the latter stage is preserved by Tsou.

The earlier stage is illustrated in (398) below. For the sake of simplicity, the c-structure does not make the distinction between clitics and non-clitics, so only the c-structure of non-clitics is presented here.
The earlier stage shown in (398) is like what is observed in Squliq Atayal, where SUBJ in f-structure corresponds to $\text{CTTOP}_\pi$ in i-structure and to PAT as well as other thematic functions in a-structure, and OBL in f-structure corresponds to $\text{CTTOP}_\vartheta$ as well as other
discourse functions in i-structure and to ACT in a-structure. In order to maintain the
one-to-one default mapping between SUBJ and CTTOPθ, which may originate from the
syntactization of GTOP to grammatical subject in a much earlier period, the conversion
function Q is applied to provide the flexibility of information insertion into i-structure
from discourse-pragmatic contexts, with the side effect that it gradually encroach on this
default mapping in the long term, a phenomenon discussed in Section 9.4.2.

On the other hand, although the OBL in f-structure maps onto several discourse
functions in i-structure, their frequency of either type or token is declining so that
CTTOPθ obtains the vast majority of tokens (Cf. Chapter 6). This phenomenon is
attributed to two mutually dependent factors. First, the agent-oriented nature of
discourse enhances the possibility that oblique actors convey given information in
information structure. Second, the precedence of oblique actors over the nominative
subjects in the linear order reduces the possibility to violate the Given-First constraint
that given information comes first. High topicality activates the agenthood of oblique
actors and heightens their effect in syntax by acquiring anaphoric prominence and
working as the addressee of imperatives. Put in the LFG terminology, the strengthened
correspondence between OBL in f-structure, ACT in a-structure, and CTTOPθ in
i-structure leads to the emergence of split-subjecthood. It explains why some oblique
arguments have the potential for the acquisition of subjecthood while others don’t. For
example, the oblique arguments in English passive constructions have no subject
properties because they are not highly topical in the absence of the interplay between
the above two factors.

The later stage of split-subjecthood is preserved in Tsou, as illustrated in (399).
Once again, only the c-structure of non-clitics is presented for convenience.
(399) Later Stage (preserved by Tsou)

Compared with the earlier stage, this stage stands for the further development of split-subjecthood. The above-mentioned factors leading to the emergence of split-subjecthood trigger off a snowball effect. With the decrease of the mapping with
discourse functions other than CTTOP, extremely high topicality turns the oblique actors in NAV clauses into mapping with CTTOP in i-structure, that is, the shift of the primary continuing topic (CTTOP) from nominative subjects to oblique actors in NAV clauses (as marked with the red color in (399)).\textsuperscript{138,139} In the meanwhile, or later on, this shift is accompanied with the broader syntactic activation of oblique actors (carrying OBL in f-structure) because the high priority which CTTOP possesses in discourse-pragmatic coherence gradually permeates into syntax. The OBL corresponding to CTTOP ends up becoming the prominent grammatical function \( \widehat{\text{GF}} \) (marked with the purple color in (399)). Although \( \widehat{\text{GF}} \) is a notion borrowed from Falk (2006), as discussed above, its existence is not manifested in all cases of split-subjecthood, which is the discrepancy between the proposal of the present thesis and that of Falk (2006). In other words, the oblique actors in NAV clauses are reanalyzed from noncore arguments to core arguments.

The diachronic development that the oblique actors in NAV clauses no longer corresponds to OBL but to \( \widehat{\text{GF}} \) after its long mapping with CTTOP may be formalized in the tendency (400).

\begin{itemize}
  \item \textbf{(400) Tendency toward Prominence Realignment}
  \item Given \( \alpha \) is the most prominent function in a representation \( p \), \( \beta \) is the most prominent function in a representation \( q \), and \( \gamma \) is the most prominent function in a representation \( r \), \( p \neq q \neq r \), for any lexical item \( x \), \( x = \alpha(\text{PRED}) = \beta(\text{PRED}) \), then \( x = \gamma(\text{PRED}) \) eventually.
\end{itemize}

Teleological explanation, such as the above tendency, regards change as aiming at a state preferable to the initial one, but it does not explain how the language reaches that

\textsuperscript{138} The directionality of language change is termed as “(linguistic) drift” in Sapir (1921). This term is not adopted in the present thesis because it implies that a certain language change follows only one possible way of development—that is, the notion of unidirectionality.

\textsuperscript{139} Put in the OLF-OT convention, the i-structural ranking gradually changes from the priority of placing important in the sentence-final position to the priority of placing given information in the sentence-middle position (i.e. Aboutness-Last >> Given-First to Given-First >> Aboutness-Last) (Cf. Chapter 7).
stage (Faarlund 1990). In addition, the above tendency is not the only driving force behind the acquisition of subjecthood. As pointed out in Cole et al. (1980), some Austronesian languages stay at the halfway stage toward the complete acquisition of subjecthood, which is also exemplified by Squiliq Atayal and Tsou. Why? The reason is simple. Once the oblique actors in NAV clauses acquire all of the morphosyntactic subject properties, the distinction between AV clauses and NAV clauses will disappear. Since language is a means of human communication, it definitely tends to maintain the diversity of expression, as stated in the following tendency:

(401) **Tendency toward Diversity of Expression**
Given a construction A is morphosyntactically distinct from construction B, A will tend to be distinct from B either semantically or discourse-pragmatically.

The second tendency is discussed in various functionalist studies, such as Goldberg (1995) and Haiman (1985). It stops the oblique actors in NAV clauses in Squiliq Atayal and Tsou from becoming real grammatical subjects, so as to make a distinction between AV clauses and NAV clauses at the levels of syntax and discourse-pragmatics.

### 9.3.3 Evidence

In order to verify the diachronic development of split-subjecthood proposed by the present thesis, it is necessary to justify two essential parts of this hypothesis. One is its premise that the oblique actor in NAV clauses in Tsou corresponds to CTTOP in i-structure. The other concerns the direction of CTTOP shift. That is to say, the discourse function CTTOP shifts from nominative subjects to oblique actors rather than from oblique actors to nominative subjects in NAV clauses in history. In what follows, Section 9.3.3.1 and 9.3.3.2, these two parts will be justified with both qualitative and quantitative evidence, even without direct evidence from historical documentation.
9.3.3.1 For $\text{CTTOP}_{\pi}$ in Tsou

The first evidence for the oblique actor as $\text{CTTOP}_{\pi}$ in Tsou comes from S. Huang (2002), which adopted Givón’s method to measure the topicality of core arguments. As elaborated in Chapter 5, topicality is measured by RD and TP. The former stands for given: the shorter in RD a nominal has, the more given its information is. The latter stands for aboutness: the more repetitions in TP a nominal has, the more important in subsequence discourse its information is. The same quantitative measurement lays the groundwork for our comparisons between Squliiq Atayal and Tsou. In addition, the statistical numbers cited below are all from narratives so that the genre difference in our comparisons can be reduced to some extent. S. Huang (2002) grouped data into three categories: the three categories of RD are high (RD<2), medium (RD= 2~10) and low (RD>10), and those of TP are high (TP>=3), medium (TP=2) and low (TP<1). In order to make comparisons possible and to highlight the divergence between Squliiq Atayal and Tsou, his grouping is modified from three categories into two: for givenness, high RD (RD $\leq$ 10) and low RD (RD > 10), and for aboutness, high TP (TP $\geq$ 3) and low TP (TP < 2). The results of comparisons are presented in Table 9-2 and Table 9-3.

Table 9-2 RD Comparison: Squliiq Atayal & Tsou

<table>
<thead>
<tr>
<th>NAV</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RD $\leq$ 10 (%)</td>
<td>RD &gt; 10 (%)</td>
</tr>
<tr>
<td>Squliiq Atayal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominative Subject</td>
<td>86.3</td>
<td>13.7</td>
</tr>
<tr>
<td>Oblique Actor</td>
<td>91.3</td>
<td>8.1</td>
</tr>
<tr>
<td>Tsou</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominative Subject</td>
<td>81.7</td>
<td>18.2</td>
</tr>
<tr>
<td>Oblique Actor</td>
<td>100.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
In Table 9-2, the RD values of nominative subject and oblique actor in NAV clauses in these two languages are put into comparison side by side. In Squliq Atayal, 86.3% of nominative subject are placed in High RD while only 13.7% of them are placed in Low RD, and 91.3% of oblique actors are placed in High RD while only 8.1% of them are placed in Low RD. This distribution has been thoroughly discussed in Chapter 5 & 6. Briefly speaking, although oblique actor is more given than nominative subject at the level of discourse-pragmatics, what really corresponds to CTTOP in i-structure is nominative subject at the level of syntax. With regard to Tsou, 81.7% of nominative subjects are placed in High RD while 18.2% of them are in Low RD, and 100% of oblique actors are placed in High RD while none of them are placed in Low RD. We may get a first impression that Tsou is like Squliq Atayal since nominative subject is less given than oblique actor as well. However, if we make a cross-linguistic comparison, we can discover that nominative subject in Tsou (81.7%) has lower percentage than that in Squliq Atayal (86.3%) while oblique actor in Tsou (100.0%) has higher percentage than that in Squliq Atayal (91.3%). In other words, the oblique actor in Tsou is much more given than the oblique actor in Squliq Atayal while the nominative subject in Tsou are less given than the nominative subject in Squliq Atayal. The discrepancy between them is even up to around 9% with regard oblique actor.

Table 9-3 TP Comparison: Squliq Atayal & Tsou

<table>
<thead>
<tr>
<th>NAV</th>
<th>High TP ≥ 3 (%)</th>
<th>Low TP &lt; 2 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squliq Atayal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominative Subject</td>
<td>28.5</td>
<td>71.5</td>
</tr>
<tr>
<td>Oblique Actor</td>
<td>46.0</td>
<td>54.0</td>
</tr>
<tr>
<td>Tsou</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominative Subject</td>
<td>31.8</td>
<td>68.1</td>
</tr>
<tr>
<td>Oblique Actor</td>
<td>60.1</td>
<td>39.9</td>
</tr>
</tbody>
</table>
On the other hand, what Table 9-3 presents is the comparison on TP. In Squilq Atayal, 28.5% of nominative subjects are placed in High TP while 71.5% of them are placed in Low TP, and 46.0% of oblique actors are placed in High TP while 54.0% of them are placed in Low TP. Nominative subject’s lower proportion of High TP than oblique actor is due to the application of the conversion function in i-structure. With regard to Tsou, 31.8% of nominative subjects are placed in High TP while 68.1% of them are placed in Low TP, and 60.1% of oblique actors are placed in High TP while 39.9% of them of are placed in Low TP. In terms of nominative subjects, their distribution in Tsou has almost no difference from that in Squilq Atayal, both with around 30% in High TP while with around 70% in Low TP. The discrepancy is under 2%. What really makes a great difference between these two languages is oblique actor. Oblique actors in Tsou (60.1%) have much higher percentage than those in Squilq Atayal (46.0%), with a gap up to 14.1%. In other words, oblique actors in Tsou are more important to subsequent discourse than oblique actors in Squilq Atayal since the referents which they refer to last longer in the subsequent information flow.

The second evidence for oblique actor as CTTOP in Tsou is qualitative, coming from the discourse study. According to H. Huang (2010), discourse organization in Tsou is oblique-actor-oriented, which is exemplified below (from H. Huang (2010:142)).

(402) panto moso la cih ci mo no mo toa
exist Aux.AV Hab one Rel Aux.AV Fil Aux.AV AV,pick
nasi ci eatatiskova.
pear Rel man
‘There was one…one man who was picking pears.’
This text is composed of three sentences. The first sentence is an existential construction introducing a new participant into the story, that is, *cihi ci mo nomo toa nasi ci eatatiskova* ‘one man who was picking pears’. Then, in the subsequent two sentences, whenever this participant is mentioned, it is always substantiated as the pronominal actor *-si*, as shown in the second and third sentences. The coherence between them is indicated by the boldface in both the text and the translations. This coherence pattern, where pronominal oblique actors play an important role in connecting previous information and expanding subsequent one, is prevailing in this language (Hui-ju Huang, p. c.). It is unlike what is discovered about Squliq Atayal in Chapter 5 and 6, where discourse coherence involves more complicated factors and nominative subjects are basically CTTOPs. It is proved again that oblique actors correspond to CTTOPs in Tsou.

To sum up, the quantitative evidence presented in the beginning of this section, both RD and TP, indicates that the oblique actor in Tsou is more topical than the oblique

---

140 H. Huang (2010) termed these oblique pronominal clitics as “genitive pronouns” since genitive pronouns and oblique pronouns are homophones.
actor in Squliq Atayal at the level of discourse-pragmatics. Oblique actor in Tsou is even more dominant in the comparison with other arguments. It even admits of no exception in RD. In addition, the qualitative evidence also demonstrates that oblique actors are the arguments whose primary discourse-pragmatic function is to connect their present sentence with both previous and subsequent discourse. With the proof of these two types of evidence, there is no denying that in Tsou oblique actors correspond to CT TOPπ in i-structure. That is to say, the premise of the diachronic hypothesis proposed by the present thesis does stand. Tsou and Squliq Atayal are divergent in which argument in f-structure corresponds to CT TOPπ in i-structure with regard to NAV clauses: oblique actors in the former but nominative subjects in the latter.

9.3.3.2 For Shift Direction

Once the premise of the diachronic hypothesis has been proved to stand, we have to move on to the evidence for the shift direction of CT TOPπ in the diachronic development of split-subjecthood. Owing to the lack of direct evidence from historical documentation, the evidence presented in this section only can be viewed as indirect one helping prove the diachronic path of split-subjecthood.

The first evidence comes from discourse innovation in Squliq Atayal. Recall that topicalization in this language does not permit resumptive pronouns in the position of core arguments (Cf. Chapter 7). 141 This has been attested in the grammaticality-judgment tests, as illustrated in (405) and (406), where the resumptive pronouns, both the nominative subject hiya’ in (405) and the oblique actor -nya’ in (406), cannot be co-referential with the GTOP watan.

---

141 Resumptive pronouns consist of two types, syntactically active resumptive pronouns and syntactically inactive ones (Asudeh 2012). The former are not island-sensitive while the latter are gap-like. Squliq Atayal rarely makes use of resumptive pronouns. They exist in the relative clauses for emphasis and not island-sensitive, but relativization are island-insensitive even without resumption (Liu 2004b).
However, the neat outcome of grammaticality-judgment tests runs afoul of what is found in the naturalistic data, where an oblique actor can be co-referential to GTOPs, serving as a resumptive pronoun, as exemplified in (407).

(407) bnkis=myan, raral, nanu’ a, ptnux-un=\textit{nha}'i cikonki qu’ toki zipun.
elder=1PGen past Par Par behead-PV=3PObl Cikonki Nom officer Japanese
‘(Speaking of) our forefathers, well, a Japanese officer was killed by them at Cikonki.’ <N>

In this example, the oblique actor \textit{nha}' ‘they’ refers to the GTOP \textit{bnkis-mayn raral} ‘our forefathers’. It is definitely not a slip of tongue since there are three tokens in our data. Conversely, there is no token containing resumptive pronominal subject co-referential to topic phrases in our data.

If the example (407) is isolated from its discourse-pragmatic context and examined under grammaticality-judgment tests, it becomes ungrammatical for informants again. Once again, we face the dilemma of two linguistic research methods. In the tradition of syntax, grammaticality-judgment tests tell, and we should ignore the exceptions found in naturalistic data; in contrast, in the tradition of discourse analysis, what really matters is the phenomenon discovered from naturalistic data, not the result of grammaticality-judgment tests. Our interpretation of this kind of examples is that they are discourse-pragmatic innovation, which just enters the language-particular grammar.
through some specific discourse-pragmatic contexts under the instant speech pressure as an unstable component not perceived by native speakers. The occurrence of the above example is attributed to the spontaneous application of the correspondence between oblique actors and CTTOP, which already exists in the general mapping of the language-particular grammar, but with the negligence on the distinction between CTTOP_{π} and CTTOP_{θ}. That is to say, it is merely a makeshift device under the instant speech pressure. When a makeshift device is repeatedly employed in discourse for a long while, it will no longer be an unstable linguistic component and start to be part of the stable language-particular grammar, as mentioned in the diachronic hypothesis proposed by the present thesis. Therefore, the discourse-pragmatic innovation in Squliq Atayal provides us the clue to the diachronic development of the CTTOP_{π} shift from nominative subjects to oblique actors in NAV clauses.

The second evidence for the shift direction is quantitative. In our diachronic hypothesis, one of the motivations behind the development of split-subjecthood is the agent-oriented nature of discourse, which is further relevant to our action-oriented perception, and manifested in the first language acquisition (Arbib 2012, Brown 1976, Hsu 1996, among others). It has to be emphasized here that the agent-oriented nature of discourse does not guarantee that agent must correspond to CTTOP_{π} in a language. Therefore, what it really means is that discourse tends to put emphasis upon agent since the perception of action is inherent in human cognition, so agent is the most common candidate for CTTOP. If the diachronic hypothesis presented in this chapter is correct—that is, agents stimulate the shift of CTTOP_{π} from nominative subjects to oblique actors in NAV clauses—, it may be predicted that agentive actors are more topical than non-agentive actors and subjects. In fact, this prediction is born out, as demonstrated in Table 9-4.
Among 235 tokens of oblique actors, 190 of them (80.9%) are agentive while only 45 of them (19.1%) are non-agentive. Agentive actors are the majority and that is why they have the capability to determine the development of oblique actors as a whole. Table 9-4 makes a comparison between agentive actors, non-agentive actors, and nominative subjects with [±GIV] and [±ABT]. As discussed in Chapter 6, the category [±GIV, +ABT] equals CTTOP in i-structure. In Table 9-4, 75.8% of agentive actors, 68.9% of non-agentive actors and 53.8% of nominative subjects correspond to CTTOP in i-structure, among which agentive actors have the highest percentage. The high percentage of agentive actors as CTTOP provides us with the clue that agentive actors will continue to obtain topicality via the repeated use in discourse-pragmatics and eventually establish the one-to-one correspondence with CTTOPπ probably after their percentage is up to 100%. In contrast, there is no way to get any evidence from the above phenomenon to prove the opposite direction of CTTOPπ shift from oblique actors to nominative subjects. If it were so, it would be impossible to explain why oblique actors lose their default mapping with CTTOPπ under the percentage up to around 70% whereas nominative subjects still maintain their default mapping with CTTOPπ only under the percentage up to 53.8%. Recall that the question-answer tests show that
nominative subjects should be CTTOP but oblique actors do not have this restriction in NAV clauses (Cf. Chapter 6).

The last evidence comes from typological research. It is not unusual that oblique argument acquires subjecthood in languages of different families, as presented in Cole et al. (1980). This phenomenon is thoroughly studied in Germanic languages because they have reliable historical documentation. Unlike what is observed in Austronesian languages, it only occurs to the subjects without nominative case, that is, oblique subject. In fact, they may not even be real grammatical subjects because some of oblique subjects in literature do not have any subject properties, but they thematically serve as agents, experiencers, and so forth—or their thematic role can be placed under the broad term “actor” like what is commonly used in the studies on Austronesian languages. Take Icelandic for instance. Oblique subjects in Modern Icelandic have acquired the syntactic subject properties, able to trigger control and raising (Andrews 1990). By contrast, the oblique subjects in Old Icelandic only manifest the incipience of syntactic subject properties. In line with Faarlund’s (1990) research on Scandinavian languages, Faarlund (2001) claims that oblique subjects in Modern Icelandic move to the IP-specifier position, which has been specialized for grammatical subjects through the transition from a non-configurational language (Old Icelandic) to a configurational language (Modern Icelandic). Faarlund (1985, 1990) attributed the specialization of the subject position to the reanalysis of the obligatorily-filled GTOP position in Old Norse and Old Icelandic, which had resulted from the grammaticalization of pragmatics. In other words, the position which oblique subjects move to is reanalyzed from the GTOP position to a specialized subject position, so oblique subjects gradually acquire subject

---

142 It is usually discussed in the functionalist literature, such as Givón (1979a) among others, that topic is syntactitized/grammaticalized as subject over time. The linguistic studies under the framework of MGG, such as Faarlund (2004) and Ledgeway (2012), often attest that GTOPs or the left peripheries are reanalyzed as subjects in history. However, what discourse functions the topics/GTOPs/left peripheries carry is always overlooked, which brings obstacles to the comparison between studies and to the further generalization of syntactic change.
properties. Put in the LFG terminology, because oblique subjects thematically serve as actors and always occupied the GTOP position perhaps with high topicality in the past, they were gradually activated at the level of syntax. The path of diachronic development and the motivation behind it are exactly the same as what are proposed for the oblique actors in Squilq Atayal and Tsou. The typological evidence once again invalidates the hypothesis proposed in Shibatani (1988b) that oblique actors lost subject properties in the diachronic development.

In addition, it has been attested that discourse-pragmatics triggers the change on grammatical function. In Old English, the recipient/indirect object of a ditransitive verb could not be passivized as its equivalent in Present-Day English in (408) (Denison 1993, Mitchell 1985a). As argued in Allen (1995, 2001), this restriction had continued even after the loss of the case distinction between recipient and theme, until the linear order between recipient and theme was regularized as recipient-theme due to the discourse/semantic property that recipient refers to human with higher topicality than theme. Example (409) is the first unambiguous example of recipient/indirect passive, which manifests that from around the fourth quarter of the 14th century, the recipient started to serve as OBJ rather secondary object (OBJθ).

(408) Mary was given presents (by Tom). (Denison 1993:103)

(409) Item as for the Parke, she is alowyd Every yere a dere and XX Coupull of Conyes and all fewell Wode to her necessarye…
‘Item: as for the park, she is allowed a deer every year and 20 pairs of rabbits and all firewood necessary to her…’
(Award Blount p.207) (Allen 2001:51)

Topicality has the capability to change the grammatical function of an argument to activate it syntactically, which is verified in English and also happens to
split-subjecthood in Austronesian languages as well.

9.4 Further Discussion

This section will delve into two issues associated with the diachronic development of split-subjecthood. One is about the nature of the grammatical function $\mathcal{GF}$, and the other is the decline of default mapping.

9.4.1 Examining $\mathcal{GF}$

As noted in passing in Section 9.3.2, contrary to the claim in Falk (2006) that the grammatical function $\mathcal{GF}$ exists in all languages with split-subjecthood, the present thesis proposes that the grammatical function of oblique actors in Squiliq Atayal is not $\mathcal{GF}$, but $\text{OBL}$. If it is, it will be unexplainable why the split-subjecthood phenomena it displays in Squiliq Atayal are much fewer than those it displays in Tsou. That is to say, only the oblique actors in Tsou are $\mathcal{GF}$ in f-structure because they have the topmost prominence in both a-structure and i-structure, which is related with the diachronic development that $\mathcal{GF}$ is a core grammatical function emerging from the composite of $\text{ACT}$ in a-structure and $\text{CTTOP}_\pi$ in i-structure. The syntactic properties of $\mathcal{GF}$ are manifested in its capability of serving as the addressee of imperatives, the prominent reflexive binder, and the controller/controllee of control constructions. In addition, it may trigger verb agreement and switch reference in morphosyntax, as described in Falk (2006), if these two morphological systems exist in a language. Recent studies show that morphological marking is more sensitive to information structure than linguists thought before (Dalrymple & Nikolaeva 2011, de Hoop & de Swart 2009). As long a oblique actors correspond to $\mathcal{GF}$ in f-structure and $\text{CTTOP}_\pi$ in i-structure, their significance in syntax and discourse-pragmatics eventually leads to special marking in morphosyntax. However, keep in mind that morphology does not always synchronize
with syntax. They are merely supplementary properties.

If oblique actors in Squiliq Atayal carry OBL in f-structure instead of $GF^*$ in f-structure, how can we make an account of their split-subjecthood phenomena about the addressees of imperatives and anaphoric prominence? First of all, being the addressee of imperatives is closely related with agenthood.\(^{143}\) Take the example of English. An imperative verb is always in the active voice. The fact that a passive imperative verb is rare is connected with the “understandable incongruity of combining an imperative with a stative non-agentive verb (Quirk et al. 1972:402) ”. Even in a passive imperative, as discussed in Huddleston (2002), the imperative itself assigns the agent role to the subject through its construction, as exemplified in (410), or the imperative is no longer a real imperative since its meaning is conventionally extended to the expression of wishes, as exemplified in (411).

(410) Be warned! (Heed this warning!) (Huddleston 2002:933)

(411) Get well soon. (Huddleston 2002:933)

The above examples manifest that even in English, a language generally without split-subjecthood, the addressee of imperatives must be an agent no matter in what voice the imperative is. The essence of imperatives is to give a requirement/command to the addressee (sometimes including the addresser) present at the speech context, in which the entity of the addressee is not only the shared information between interlocutors but also the important information to the subsequent discourse. That is to say, the entity of the addressee carries CTTOP in i-structure without the distinction between $\text{CTTOP}_a$ and $\text{CTTOP}_o$. Therefore, as long as an argument has the status of being

\(^{143}\) The covert argument of imperatives has a dual nature as being an agent and an addressee simultaneously. However, these two roles are separated because the covert argument of the imperatives in Dogon languages is unable to bind reflexives, a typical syntactic property of agenthood (Heath & Dyachkov 2015).
both agent and continuing topic, it can function as the addressee of imperatives. In English this argument refers to the grammatical subject in the active voice and rarely in the passive voice, whereas in Squliq Atayal it refers to the grammatical subject in AV and the oblique actor in NAV. Why does English not have the oblique in the passive voice as the addressee of imperatives? It is due to the fact that the oblique in the passive voice in English conveys new information, as pointed out in Ward, Birner & Huddleston (2002), in conflict with the requirement of the addressee of imperatives.

Likewise, having anaphoric prominence does not mean that the argument is $\mathcal{GF}$ in f-structure. Binding is a complex linguistic issue since it is related not only with relational and thematic hierarchies but also with the precedence in linear order (Bresnan 2001, Dalrymple 1993, Jackendoff 1972, among others). That is to say, the three major modules in LFG, f-structure, a-structure and c-structure, are all involved. Section 9.2.2.2 shows that a reflexive pronoun as the nominative subject in a NAV clause can be bound by the oblique actor but this binding relation cannot be reversed. Hence, there is no doubt that Squliq Atayal follows the following condition:

(412) **Thematic Hierarchy Condition on Anaphora** (Falk 2001:175)

$A$ binds $B$ iff $A$ and $B$ are coindexed and $A$ is higher than $B$ on the thematic hierarchy.

Manning (1996a) even proposed that a-structure is the locus of anaphoric binding. Thus, the posterior quantifier *nanak* ‘only/self’ in Squliq Atayal is lexically specified with the constraint (413).

(413) *nanak* ‘self’

$$
((\mathcal{GF}^* \mathcal{GF} \uparrow) \mathcal{GF}_a(\hat{\theta}) \mathcal{INDEX}) = (\uparrow \mathcal{INDEX}) \\
\neg (\rightarrow \mathcal{TENSE})
$$
It demands that the coreferential index of a reflexive pronoun (↑ INDEX) equal that of a grammatical function which corresponds to the prominent thematic function (GFα(θ) INDEX) and that the antecedent exist with the reflexive pronoun in the same finite clause ¬ (→ TENSE). The confinement to the same finite clause is based on the fact that long-distance reflexive binding is not allowed, as presented in the contrast between (414) and (415). The reflexive pronoun hiya’nanak ‘himself’ is bound to the NAV actor tali’ in the same finite clause in the former whereas it is bound the NAV actor sayun in other finite clause in the latter. Only the former is grammatical.

(414) baq-un na’ sayun, qu’ [bhy-an na’ tali’, qu’ hiya’, nanak].
know-PV Obl Sayun Nom hit-LV Obl Tali Nom 3SNeu self
‘Sayun knows that Tali hit himself.’ <E>

(415) *baq-un na’ sayun, qu’ [bhy-an na’ tali’, qu’ hiya’, nanak].
know-PV Obl Sayun Nom hit-LV Obl Tali Nom 3SNeu self
Intended for ‘Sayun knows that Tali hit herself.’ <E>

Although Falk’s (2006) proposal is a brilliant idea in that GF in f-structure lumps together the thematic hierarchy condition and the relational hierarchy condition, the existence of GF seems unnecessary to this language. So far, the only two split-subjecthood phenomena in Squiliq Atayal can be well-explained through a-structure. Remind that reflexive binding can be viewed as one part of discourse coherence so it is definitely related with i-structure.

In the proposal in the present thesis, what really differentiates Tsou from Squiliq Atayal in terms of the grammatical function which oblique actor has in f-structure is the controller of control constructions. Control constructions are similar to the addressee of imperatives and anaphoric prominence in two aspects: on the one hand, they contain a phonetically-unrealized argument, like imperatives; on the other hand, the reference of
the phonetically-realized argument is within the linguistic context, not in the extralinguistic one, like anaphoric prominence. Nevertheless, in addition to the influences from a-structure and i-structure, control constructions are strongly restricted by f-structure. As stated in Falk (2001), “[t]he controller must be a core function because noncore functions are essentially grammaticalized thematic relations, and cannot have relations to other predicates (p. 137).” On the other hand, the grammatical subject of a nonfinite open complement clause cannot be controlled by an oblique actor since as a noncore argument its existence is highly confined by lexical semantics. This is reflected in the lexical rule (416), which requires that the grammatical subject of an open complement clause should be controlled by a core grammatical function in the upper clause.

(416) **Functional Control Rule** (Falk 2001:137)

If (↑ XCOMP) is present in a lexical form, add the equation:

(↑ CF) = (↑ XCOMP SUBJ).

For some languages, like English, the specific instances of the core grammatical function vary from one verb to another, being either subjects or objects. For some languages, like Tsou and Squiliq Atayal, the core grammatical function tends to be a fixed one.

The case of control in Tsou is shown in the following example, repeated from Section 9.2.2.1.

(379) i-’o ahiy-a pa-pon-i na taini. (M. Chang & Tsai 2001:9)

NAV-1S.Gen force-PV Cau-eat-AV Nom 3S.Nom

‘I forced him to eat.’

The grammatical subject of the nonfinite verb *paboni* ‘eat’ is controlled by the oblique
actor - 'o ‘by me/my' instead of the nominative subject taini ‘he’. It proves that oblique actor in Tsou is a core grammatical function in essence, not a noncore grammatical function such as oblique. In other words, the existence of  is obligatory to predicates in Tsou.

The f-structure of this example is presented in (417). The red line indicates the control between the oblique actor and the grammatical subject of the open complement clause. The left invisible object of the open complement clause can only be controlled by the other core argument, nominative subject, as indicated in the blue line. There is a default control constraint in Tsou, as show in (418), which regularizes control constructions in general.

(417)

```
PRED 'ahiya 'force (SUBJ, , XCOMP)'
VOICE NAV
SUBJ taini 'he'
GF 'o 'me'
XCOMP [PRED 'cause-eat (SUBJ, OBJ)'
VOICE AV
SUBJ OBJ]
```

(418) **Default Control Constraint in Tsou**

\( ( \uparrow \text{GF}) = ( \uparrow \text{XCOMP SUBJ}) \)

As for the case of control in Squilq Atayal, it is illustrated in the following example, repeated from Section 9.2.2.2.
The grammatical subject of the nonfinite verb *mihiy* ‘hit’ is controlled by the nominative subject *sayun* instead of the oblique actor *tali*. This fact manifests that the oblique actor in Squiliq Atayal is definitely not a core grammatical function in f-structure. Its grammatical function is still what has been claimed, oblique (OBL).

The f-structure of this example is presented in (419), where the red line indicates the control between the nominative subject and the subject of the open complement clause. This language has a default control constraint, as show in (420).

(419) **Default Control Constraint in Squiliq Atayal**

\[
\begin{array}{l}
\text{PRED} & qrqan 'stop (SUBJ, OBL, XCOMP)'
\\
\text{VOICE} & \text{NAV}
\\
\text{SUBJ} & ["Sayun"]
\\
\text{OBL} & ["Tali"]
\\
\text{XCOMP} & \begin{array}{l}
\text{PRED} & \text{mihiy 'hit (SUBJ, OBJ)'}
\\
\text{VOICE} & \text{AV}
\\
\text{SUBJ} & \\
\text{OBJ} & \text{kuzing 'me'}
\end{array}
\end{array}
\]

(420) **Default Control Constraint in Squiliq Atayal**

\[
(↑ \text{SUBJ}) = (↑ \text{XCOMP SUBJ})
\]

So far, we have demonstrated the relationship between the controller of control constructions and \( \mathcal{G}_F \). Recall that \( \text{CTTOP}_x \) in i-structure is the most important transitional point in information chaining. It follows up the prior discourse topic and passes it to the subsequent discourse line. Once we realize this, it is not difficult to imagine that an
argument corresponding to \textit{CTTOP}_x in i-structure diachronically becomes \textit{GF} in f-structure, obtaining the status of being a default controller determining the referent of the invisible argument within a subcategorized open complement clause. In the same vein, it is also probable that an argument corresponding to \textit{CTTOP}_x in i-structure diachronically obtains the status of being a default controllee, which plays an essential role in following up the information presented in the main predicate. Again, this argument still carries the grammatical function \textit{GF} in f-structure. Based on typological evidence, Falk (2006) hypothesizes that languages in which the controllee is \textit{GF} employ anaphoric control exclusively, while languages in which the controllee is \textit{PIV/SUBJ} employ functional control. Generally speaking, the languages which have a non-subject core argument as the default controller/controllee of control constructions must have the grammatical function \textit{GF} as an independent function wielding its influence in syntax. For other languages, the syntactic influence of \textit{GF} is incorporated into \textit{SUBJ} in f-structure. Being the addressee of imperatives and having anaphoric prominence are merely supplementary syntactic properties. As for verb agreement, it does not attest the existence of \textit{GF} since Tsou has two sets of verb agreement: thematic agreement, which marks the thematic role of grammatical subject, and person/number agreement, which originates from clitics and marks the person and number of actor. Which verb agreement really matters to f-structure is uncertain. Switch reference does not exist in Tsou so it can only be viewed as a supplementary property. Supplementary properties are not exclusive to \textit{GF}. Therefore, they may serve collectively as the evidence for the existence of independent \textit{GF} in a language but not individually.\textsuperscript{144} As for the topicalization of oblique actor in Tsou, it may show that \textit{GF} is under \textit{SUBJ} in the

\textsuperscript{144} Some studies propose that the boundary between two grammatical functions is not a clear-cut line but a continuum (De Wolf 1988, Edwards 2012a, b, Shibatani 1988b, among others). As demonstrated in this chapter, the grammatical function of NAV actor seems to form a continuum between OBL in Squliq Atayal and \textit{GF} in Tsou. The synchronic continuum may be a substantiation of diachronic development. Since the continuum view has not formed a complete system, we take a more conservative position in this chapter.
relational hierarchy and that the extension of topicalization to oblique actor indicates the application of topicalization in this language has been expanded under unknown motivation so far.

As discussed in previous LFG literature, grammatical functions are decomposed into binary features in LMT (Cf. Chapter 3). The feature \([\pm r]\) stands for whether or not the mapping with thematic functions is restricted and the feature \([\pm o]\) stands for whether or not a grammatical function possesses objecthood (Bresnan et al. 2016). In terms of these two features, the grammatical function \(\hat{GF}\) is the same as OBL, with the positive value of \([\pm r]\) and the negative value in \([\pm o]\), since both of them exclusively correspond to ACT in a-structure and possess no objecthood, as illustrated below:

\[
\begin{array}{cccc}
\text{OBL} & \text{GF} & \text{SUBJ} \\
\begin{array}{ccc}
[r] & [o] & [i] \\
+ & - & - \\
+ & - & + \\
- & - & + \\
\end{array}
\end{array}
\]

It is proposed here that what differentiates \(\hat{GF}\) from OBL is another binary feature: \([\pm i]\), which stands for whether or not a grammatical function plays the most important role in discourse-pragmatic coherence. Both of \(\hat{GF}\) and SUBJ have the positive value of \([\pm i]\) while OBL has the negative value of \([\pm i]\). Although \(\hat{GF}\) resembles SUBJ with regard to \([\pm o]\) and \([\pm i]\), their difference comes from whether or not they are restricted in the mapping with thematic functions. The above binary distinctive features show that the grammatical function \(\hat{GF}\) is only one feature away from either OBL or SUBJ. That is why \(\hat{GF}\) is present at the transitional stage from oblique actor to grammatical subject in the diachronic development. The argument carrying \(\hat{GF}\) in f-structure may be viewed as a real syntactically-defined “oblique subject”.

344
9.4.2 Breaking Default Mapping

The diachronic hypothesis proposed in Section 9.3.2 has little consideration of default mapping since there is no study on the results of question-answer tests in Tsou. Thus, the hypothesis is mainly about the shift of $CTTOP_\pi$ and its correlation with $GF$. In fact, the shift of $CTTOP_\pi$ and default mapping are independent from each other. In essence, default mapping has nothing to do with the distinction between $CTTOP_\pi$ and $CTTOP_\theta$ so that the shift of $CTTOP_\pi$ does not immediately lead to the establishment or breaking of default mapping.

Despite the lack of the evidence from question-answer tests, the naturalistic data from Tung (1964) provide us some clues. The sentences in (422) and (423) form a question-answer pair.

(422) ci sia na mo mosi tan'e ta oko? (Tung 1964:406-407)
who put here child
‘Why, who (is) the (one who) put the child here?’

(423) oh ken a hafa no inosi ho trongasi no mo fitfi no tu'nu.
taken its-mother left precipitous cliff
‘Its-mother had taken (the child) and left (it) on the precipitous cliff.’
(Tung 1964:406-407)

The first sentence is a question inquiring the agent of the event. Although Tung’s glosses are sketchy, the prefix $m$- of the main verb $mosi$ ‘put’ indicates that this sentence is an AV clause and that $sia$ ‘who’ is a pseudo-cLEFTed grammatical subject as well as the agent. The answer part in (423) is a coordinate clause, the first part of which answers the inquiry by showing the referent of the agent $inosi$ ‘its-mother’. The agent is the new information offered by the addressee and it is preceded by the oblique case $no$, which
indicates that this sentence is a NAV clause and that the patient oko ‘the child’ is the omitted nominative subject.\textsuperscript{145} Given information is placed on the position of nominative subjects and new information is on that of oblique actors. This at least proves one thing. That is, even though \texttt{CTTOP}_\pi corresponds to oblique actors in Tsou, it does not mean that the correspondence between them has been set default but just means that the tokens of new information conveyed by oblique actors have been reduced to such as a large extent that S. Huang’s (2002) quantitative study cannot display their existence (Cf. Section 9.3.3.1). Further studies are needed to tell whether the default mapping between \texttt{CTTOP}_\pi and nominative subject in NAV clauses still exists in Tsou.

Although the shift of \texttt{CTTOP}_\pi does not guarantee the establishment or breaking of default mapping, it is undoubted that the application of the conversion function \texttt{Ϙ}, a discourse-pragmatic makeshift, eventually breaks the default mapping between \texttt{CTTOP} and nominative subject in NAV clauses. It is proved in the limited evidence from Tagalog in Kroeger (1993) and Naylor (1975), as shown in (424), (425) and (426).

(424) Ano ang kinain mo?  (Naylor 1975:48)
    What Nom Perf.OV-eat 2SGen
    ‘What did you eat?’

(425) Kinain ko \textit{ang=}isda.  (Naylor 1975:49)
    Perf.OV-eat 1SGen Nom=fish
    ‘I ate the fish.’

(426) Kumain ako \textit{ng=}isda.  (Naylor 1975:49)
    AV.Perf-eat 1SNom Gen=fish
    ‘I ate (some) fish.’

\textsuperscript{145} Case marking system in Tsou is more complex than that in Squiliq Atayal. Details can be referred to in Zeitoun (1992, 2000).
The first sentence is a NAV clause, inquiring about the patient of the event. The interrogative word \textit{ano} ‘what’ is a pseudo-clefted nominative subject. The two sentences in (425) and (426) are two different ways to answer this question. The former places the given information \textit{ko} ‘I/my’ on the position of oblique actor and the new information \textit{isda} ‘fish’ on the position of nominative subject. This sentence is crucial in that its distributional pattern of information is unlike what we observed in Squiliq Atayal and Tsou. Nominative subjects in NAV clauses are able to convey new information in the answer part. Apparently, the default mapping between CTTOP and nominative subjects in NAV clauses has been broken in Tagalog. From the diachronic perspective, Tagalog walks further than Tsou and definitely much further than Squiliq Atayal with regard to the decline of default mapping.\footnote{This may be attested in the rough comparison of RD. The average RD of nominative subject in NAV clauses is 10.01 (Cooreman, Fox & Givón 1984; Kroeger 1993). As shown in Table 9-2, the nominative subject in NAV clauses with low RD (RD > 10) only amounts to 18.2\% (based on the data in S. Huang (2002)). Although it is difficult to compare an average with a percentage, we still can speculate that 18.2\% of low-RD tokens are unable to overwhelm 81.8\% of high-RD tokens to the extent that their average is up to 10.01 as the average in Tagalog. Despite the lack of raw data, it is quite possible to discover in statistics that the nominative subject in NAV clauses in Tagalog conveys newer information than that in Tsou.}

As for the second kind of answer in (426), it places the given information \textit{ako} ‘I’ on the position of grammatical subject and the new information \textit{isda} ‘fish’ on the position of object. Its distributional pattern of information is like the AV clauses in Squiliq Atayal and Tsou. What is intriguing, as pointed out in previous studies on Tagalog, such as Schachter & Otanes (1976) and Schachter (1976), the semantic interpretation of \textit{isda} ‘fish’ varies from its grammatical function: definiteness for subject while indefiniteness for non-subject.\footnote{De Wolf (1988) pointed out with the evidence from Tagalog texts that the nominative subject in NAV clauses is not always definite. There is no surprise. As the case of Squiliq Atayal discussed in Chapter 6, the default mapping does not mean that there is no exception in naturalistic data, which is affected by the discourse-pragmatics under the time pressure in speech.} Definite expressions indicate that the existence of their referents or quantity is presupposed and identifiable (and sometimes unique) to the addressee, which is substantiated in the use of the definite article \textit{the} in English
(Payne & Huddleston 2002). It does not matter for definite expressions whether or not the information they convey is presumably possessed by the audience in current information flow. Take an example of (425). The oblique subject *isda* ‘fish’ conveys a piece of new information to the addressee since the addressee has no idea about what the patient is; however, once the addressee receives this information, he is assumed to easily identify the referent of the nominative subject *isda* ‘fish’ with a particular fish he or she has in mind. It is highly possible that in the history of Tagalog the default mapping between CTTOP and grammatical subject gradually switched to the default mapping between definiteness and grammatical subject.\(^{148}\) In other words, the usage of grammatical subject in Tagalog may have undergone a transition from the expression of given information to the expression of an identifiable referent, a decontextualizing process from discourse-pragmatics to semantics. The common ground of the two terminals in this process is the strong tendency that given information always refers to something identifiable. The catalyzer of this process is the erosion brought about by the conversion function \(Q\).

Generally speaking, the default mapping is not directly related with the shift of CTTOP\(_e\) from nominative subject to oblique actor in NAV clauses. They are not in the causal relationship. This diachronic development can be reconstructed through the

---

\(^{148}\) In the history of Chinese, the grammatical subject seems to have undergone the same transition. Unmarked sentences in the Old Chinese of the earliest documents, Yin oracle bone inscriptions, follow the given-new discourse principle with grammatical subjects conveying given information (Kryukov 1996). Later, the demonstrative pronoun *shi* came to existence as the grammatical subject anaphoric to a previously mentioned entity, with such a frequent use that then it was gradually grammaticalized as a copular verb, as it was in Middle Chinese and it still is in Modern Chinese (Ota 2003[1958], Pulleyblank 2006[1995], Shi & Li 2001, Wang 2004[1958], Xiang 2010, Xiao 2006, Yang & He 2001, inter alia). Its frequent use in Late Old Chinese shows that there still was a default mapping between grammatical subject and given information at that time (Liu 2008). However, the default mapping between grammatical subject and definiteness in Modern Chinese seems to be gradually broken since the use of specific indefinite subject has been gradually expanded since Middle Chinese, as manifested in Liu’s (2010) study on the diachronic development of the numeral-classifier phrase *yige* ‘one’ in the subject position and in Tsai’s (2003) proposal that the existential modal verb *you* in Modern Chinese is undergoing grammaticalization towards an indefinite article in the subject position. Teng (2012[1985]) claimed that subjects in Modern Chinese still only convey given information. Frankly speaking, his examples and claim are not commonly accepted. Based on my personal intuition on Mandarin Chinese, they do not hold.
comparison between Tsou and Tagalog.

9.5 Summary

Split-subjecthood is prevalent in Austronesian languages. It is only attested in the phenomena of the addressee of imperatives and anaphoric prominence in Squiliq Atayal, while it is attested in several morphosyntactic properties in Tsou. Based on the variations between Squiliq Atayal and Tsou, the present thesis proposes a hypothesis that with regard to the diachronic development of split-subjecthood, Squiliq Atayal preserves the earlier stage while Tsou preserves the later stage. There is the CTTOPx shift from nominative subjects to oblique actors in i-structure. It further triggers the reanalysis of OBL to $\mathcal{GF}$ in f-structure, which is manifested in the fact that oblique actor serves as the default controller/controlee of control constructions. In addition, the addressee of imperatives and anaphoric prominence cannot be viewed as true split-subjecthood phenomena since they come from the interplay between $\theta$ and CTTOPx having nothing to do with the separation of $\mathcal{GF}$ from SUBJ in f-structure. Furthermore, it is demonstrated that the default mapping between the nominative subject in NAV clauses and CTTOP does not cease to exist even after the CTTOPx shift to oblique actors in Tsou. Nevertheless, the application of the conversion function $Q$ continues weakening this default mapping until it is broken as displayed in Tagalog.
10. Concluding Remarks

It is the peculiarity of knowledge that those who really thirst for it always get it.

-- Richard Jefferies

10.1 Contributions

The present thesis investigates the syntactic interactions with information structure in Squilq Atayal, a minority language of the Austronesian family in Taiwan. Adopting the LFG framework, it makes several descriptive and theoretical contributions with the support of quantitative and qualitative evidence, as summarized in the following:

<table>
<thead>
<tr>
<th>No.</th>
<th>Contributions</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Providing theoretical models of feature-based discourse functions and i-structural constraints for LFG;</td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>Incorporating one-to-many correspondence into LFG;</td>
<td>4</td>
</tr>
<tr>
<td>3.</td>
<td>Studying the topicality of core arguments, subject, object, oblique, GTOP and GFOC, in naturalistic data with a statistical method;</td>
<td>5</td>
</tr>
<tr>
<td>4.</td>
<td>Describing the information chaining in naturalistic data;</td>
<td>5</td>
</tr>
<tr>
<td>5.</td>
<td>Identifying what discourse functions in i-structure subject, object and oblique correspond to through the mapping in unmarked clauses;</td>
<td>6</td>
</tr>
<tr>
<td>6.</td>
<td>Discovering that grammatical voices in Squilq Atayal belong to both the role-remapping system and the information-salience system;</td>
<td>6</td>
</tr>
<tr>
<td>7.</td>
<td>Showing that the mapping with i-structure complements the lexically determined mapping approach.</td>
<td>6</td>
</tr>
<tr>
<td>8.</td>
<td>Identifying what discourse functions in i-structure GTOP correspond to through the mapping in marked clauses;</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Identifying what discourse functions in i-structure GFOC correspond to through the mapping in marked clauses;</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Proposing the Informational Mapping Theory;</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Providing a diachronic hypothesis to account for the emergence of split-subjecthood in Austronesian languages;</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Examining $\hat{G}F$ and default mapping with the evidence from Squiliq Atayal, Tsou and Tagalog.</td>
<td></td>
</tr>
</tbody>
</table>

Generally speaking, it is manifested in Squiliq Atayal that from the synchronic perspective, syntax and information structure are closely related with each other, no matter what structure is investigated, unmarked clauses or marked clauses, even though Squiliq Atayal is a typical configurational language. If the diachronic perspective is adopted, the change in information structure eventually affects syntax, as presented in the hypothesis about the emergence of split-subjecthood in Austronesian languages.

### 10.2 For Future Research

No academic study is impeccable. In addition to the accomplishments achieved by the present thesis, some linguistic issues are left to future research. First of all, the methodological and theoretical innovations proposed in the present thesis still need to be applied to other languages, so as to be examined on their typological validity and reliability. Secondly, the present thesis tries to incorporate the concept of gradience into theoretical syntax by formalizing several different spans of percentage. There is no doubt that this strategy of formalization is rather rough. The most ideal way should have the ability to achieve the precision of calculation displayed in fuzzy math and advanced statistics. Third, the present thesis deals only with information-structural issues on nominals, but neither of predicate-related and proposition-related issues is discussed. This is due to lack of evidence and to their complexity. Last, although the diachronic hypothesis about the emergence of split-subjecthood in Squiliq Atayal and Tsou can
never be attested by historical documentation, this hypothesis still can be examined in other languages, especially those Indo-European languages with split-subjecthood which have long-standing writing tradition. However, to my limited knowledge, the interactions between syntax and information structure are seldom touched upon in historical linguistics. Hope that the completion of the present thesis is not merely the consummation of a linguist’s wish but also the commencement of future studies on the syntactic interactions with information structure.
Appendix 1:
Effect of Different Components

This appendix attempts to examine what linguistic components, syntactic roles, semantic roles, topicality, and phonetic forms, affect the word order in Squilq Atayal and to what extent they explore their effect. By adopting the multinomial logistic regression, several models are built to analyze which one has the most optimal prediction of the word order and then analyze the effect of different component.

A1.1 Multinominal Logistic Regression

What affects word order is the enigma that syntacticians spend their lifelong time trying to solve. Different components have been proposed, such as syntactic roles, semantic roles and topicality in particular under discussion of this appendix. In addition to theoretical explanations, we are able to analyze this significant syntactic issue with the assistance of statistical methods. What we are going to employ is the multinomial logistic regression, a statistical method to build hypothetical models and evaluate the importance of different syntactic components (i.e. variables in statistics).

Logistic regression analysis is a way of predicting an outcome variable (i.e. dependent variable) that is categorical from predictor variables (i.e. independent variable) that are continuous or categorical (Gray & Kinnear 2012, Johnson 2008, Manning 2003). The model produced by logistic regression is nonlinear. When there are two categories of the outcome variable, “the outcome variable, $\hat{Y}$, is the probability of having one outcome or another based on a nonlinear function of the best linear combination of predictors (Tabachnick & Fidell 2007:438)”, as illustrated below:
In the above formulae, $\hat{Y}_i$ is the estimated probability that the case $i$ is in one of the categories from 1 to $n$, and $u$ is the linear regression equation with constant $A$, coefficients $B_j$, and predictors $X_j$, in which $j$ is from 1 to $k$. The logarithmic transform of (427) is as the formula (428). It is easy for statisticians to map logistic values obtained from (428) to odds or probabilities under the assumption that there is a linear relationship between predictors and the logarithmic transform.

(428)

$$\log\left(\frac{\hat{Y}}{1 - \hat{Y}}\right) = A + \sum B_jX_{ij}$$

One of the subtypes of logistic regression is “multinomial logistic regression (MLOGIT),” wherein the outcome variable has more than two categories. It is assumed that, although the value of the outcome variable can fall into any of its categories, predictor variables increase the probability of developing the outcome continuously throughout the range of either variable. By using the logarithmic transformation, logistic regression calculates the proportionate change in odds by dividing the odds after a unit change in the predictor by the odds before that change (Field 2009). If this proportionate change in odds is statistically significant, that particular predictor variable makes a significant influence upon the outcome variable. In the analysis of the present appendix, the outcome variable is the word order—to be more specific, the position in a clause. The predictor variables are syntactic roles, semantic roles, topicality, and phonetic
forms. MLOGIT will calculate whether a unit/category change of these predictors, for example when the syntactic role of a nominal expression is changed from subject to object, brings about any difference on the probability of being placed in any position in a clause. In addition, MLOGIT can estimate the fitness of models built upon the combinations of different predictor variables to the observed values, which makes us able to judge which model is the optimal.

Since the outcome variable is word order, the positions of noun phrases are coded. The preverbal position is coded with 0, and the postverbal positions are divided into the sentence-medial position coded with 1, and the sentence-final position coded with 2. Null arguments are recovered into their coreferential noun phrases in the previous discourse, and their coding is based on the position of their recovered noun phrases. Take the following sentence for instance:

(429) nanu’, tayal hiya’ ga, baq karaw qhoniq Ø.
       par Atayal Emp Top AV.know AV.climb tree Ø

‘Well, (speaking of) the Atayal themselves, (they) know how to climb trees.’ <N>

In (429), the preverbal noun phrase tayal ‘the Atayal people’ is coded with the value 0. Moreover, the null argument Ø will be recovered into its coreferential noun phrase, that is tayal ‘the Atayal people’, and it is coded with 2 because it is located in the sentence-final position. As for the qhoniq ‘tree’, it is coded with 1 because it is postverbal and in the sentence-medial position.

The predictor variables consist of topicality (RD & TP), syntactic roles, semantic roles, and phonetic forms. They are the components of language with the possibility to affect our outcome variable, the word order. Their coding strategies are listed below.
(430) Coding Strategies

- **RD**: 0 = new information; 1 = more given information; 2 = less given information.
- **TP**: 0 = unimportant information; 1 = less important information; 2 = more important information.
- **Syntactic roles**: 0 = GTOP; 1 = subject; 2 = oblique; 3 = object; 4 = GFOC.
- **Semantic roles**: 0 = no; 1 = agent; 2 = patient; 3 = theme; 4 = beneficiary/instrument; 5 = location/path/source; 6 = temporal; 7 = others.
- **Phonetic forms**: 0 = monosyllabic; 1 = disyllabic; 2 = polysyllabic.

The measurement of RD and TP is identical with that in Chapter 5. The coding of syntactic roles is based on the recovered case marking of each noun phrase and pronoun since case marking of noun phrase is omissible in Squilq Atayal. Put differently, if a noun phrase is not preceded by its case marker, we will code it according to its original case marker. In the same vein, GTOPs are coded based on whether they can be followed by the topic marker *ga*. The data used in this study are the same with those in Chapter 5 as well. However, we only examine transitive clauses, removing intransitive sentences and sentences only containing a nominal predicate.

In order to estimate the effect of the above predictor variables, we design the following various models.

(431) Models of MLOGIT Analysis

- **Model 1**: RD and TP
- **Model 2**: phonetic forms
- **Model 3**: semantic roles
- **Model 4**: RD, TP, semantic roles, and phonetic forms
- **Model 5**: syntactic roles, RD, TP, semantic roles, and phonetic forms.

Each of the above models contains different combinations of predictor variables. The first four models (Model 1-4) are called “deficient models” for convenience, and the last one (Model 5) is the “full model” since only the last model contains all of the predictor
variables. All of these models will undergo the MLOGIT analysis, but only the statistical report of the full model will be presented since it makes an overall comparison of all predictor variables. Deficient models are designed to highlight the insufficiency of a single component or some possible combinations of several components.

A1.2 Deficient Models

Deficient models contain only one or several, but not all, predictor variables. Their outcome variable is the same: word order. MLOGIT will estimate the accuracy of their predictions, which is presented in the following subsections.

A1.2.1 Model 1, 2 & 3

Model 1 has two predictor variables, RD and TP, both of which are associated with topicality. When only topicality is taken into consideration, the overall accuracy of this model is up to 54.8%, as shown in Table 0-1. In other words, topicality of nominal expressions can only correctly predict around a half of their positions. What is problematic is that topicality cannot make any prediction of the occurrence of preverbal elements (i.e. GTOPs and GFOCs), the accuracy of which is 0.0%. The insufficiency of Model 1 points out that, although there is no denying that some fundamental relations like word order may have arisen from discourse-pragmatic properties (Bybee 2010, Du Bois 1987, Haiman 1985, among others), discourse-pragmatic properties alone cannot determine all of syntactic facts discovered without the cooperation of other components. Model 1 proves again that topicality is not the only determinant of word order inSquliq Atayal.
Table 0-1 Accuracy of Model 1

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
</tr>
<tr>
<td>Initial</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>Middle</td>
<td>0</td>
<td>308</td>
</tr>
<tr>
<td>Final</td>
<td>0</td>
<td>160</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model 2 contains only one predictor variable—phonetic forms. They are categorized by the number of syllable of each nominal expression. Just like Model 1, the accuracy of this model is still low, only 69.4%, as shown in Table 0-2, but it is rather higher than Model 1. The most inaccurate part still comes from the sentence-initial position. It is apparent that phonetic forms have no effect on whether or not a nominal can be located in the sentence-initial position. They however raise the accuracy of our prediction of the middle and final positions, which is due to the fact that clitics are monosyllabic or disyllabic. Clitics in Squilq Atayal are adjoined to the first element in a clause, and undoubtedly they are in the sentence-medial position.

Table 0-2 Accuracy of Model 2

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
</tr>
<tr>
<td>Initial</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Middle</td>
<td>0</td>
<td>350</td>
</tr>
<tr>
<td>Final</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There is only one predictor variable in Model 3—semantic roles. Table 0-3 shows that the overall accuracy of this model is still low and only 60.1%, just like what we have seen in Model 1 and 2. However, the accuracy of the initial position rises to 43.6%, which is a great progress compared with the failure of the prediction of this position in previous models. This progress is associated with the tendency that agent and temporal always function as GTOPs as well as GFOCs. In addition, it is shown that only semantic roles do not work well on determining the word order of nominal. A grammar with only semantic roles is deficient.

Table 0-3 Accuracy of Model 3

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
</tr>
<tr>
<td>Initial</td>
<td>44</td>
<td>19</td>
</tr>
<tr>
<td>Middle</td>
<td>2</td>
<td>299</td>
</tr>
<tr>
<td>Final</td>
<td>6</td>
<td>137</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A1.2.2 Model 4

Model 4 is a model composed of all the predictor variables in previous models: RD, TP, phonetic forms, and semantic roles. As shown in Table 0-4, the overall accuracy of Model 5 (76.4%) is much higher than all of previous models. Moreover, all the correct percentages of different positions are higher as well. The interplay of components like topicality, semantic roles, and phonetic forms does affect the word order in Squilq Atayal. The MLOGIT analysis of this model also aims to demonstrate to what extent a grammar model without syntactic roles predicts the reality observed from naturalistic data. Even though some linguists propose that syntactic roles are merely restricted
neutralization of other components (Cf. Van Valin 2005, inter alia, Van Valin & LaPolla 1997), it is obvious at least in Squilq Atayal that the accuracy of a model without syntactic roles is limited.\textsuperscript{149} Any approach completely abandoning the notions of syntactic roles is untenable for Squilq Atayal. Of course, more MLOGIT studies are needed to examine those “semantically-marked” languages, the morphosyntax of which relies only upon semantic roles, not upon syntactic roles. Perhaps, syntactic roles still affect the morphosyntax of those languages to a small extent, rather than to no extent.

Table 0-4 Accuracy of Model 4

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
</tr>
<tr>
<td>Initial</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>Middle</td>
<td>3</td>
<td>382</td>
</tr>
<tr>
<td>Final</td>
<td>5</td>
<td>68</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A1.3 Full Model & Interpretation

Full model takes into account all of predictor variables like syntactic roles, semantics, topicality (RD & TP), and phonetic forms to examine whether each of them affects word order and how important to the overall pattern it is. The difference between the full model and other models lies upon the participation of syntactic roles. As we can

\textsuperscript{149} Linguists under the framework of Role and Reference Grammar reject the existence of syntactic roles as linguistic primitives, claiming that there is the Actor-Undergoer hierarchy assigning semantic roles with two macroroles, Actor and Undergoer, as shown below (LaPolla 1990:23):

\begin{center}
(i) Actor
\begin{itemize}
  \item agent
  \item effector
  \item experiencer
  \item locative
  \item theme
  \item patient
\end{itemize}
Undergoer
\end{center}
see in Table 0-5, the overall accuracy of the full model rises to 99.3%, and the accuracy of each position is above 90%. There is no doubt that syntactic roles play a prominent role in the prediction.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
</tr>
<tr>
<td>Initial</td>
<td>101</td>
<td>0</td>
</tr>
<tr>
<td>Middle</td>
<td>0</td>
<td>481</td>
</tr>
<tr>
<td>Final</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When it comes to the effect of these components in the full model, however, MLOGIT analysis tells that there is great multicollinearity among variables since the $R^2$ value is 0.97 above 0.25, as shown at the bottom of Table 0-6; that is to say, there is a great overlap among different components, which is unsurprising to linguists because they are highly interactive either in human languages. This overlap can also be illustrated by comparing the full model in Table 0-5 with the additional model containing only syntactic roles in Table 0-7. The additional model has high accuracy of prediction as well, 94.5%, which is only a little lower than that in the full model. This highly similarity falls afoul of the results we get from other models, wherein other predictor variables alone can have less than 80% of accuracy. It is problematic for the MLOGIT analysis to estimate the effect of predictor variables if there is a great overlap among them (Field 2009, Tabachnick & Fidell 2007).
Table 0-6 *Likelihood Ratio Tests*

<table>
<thead>
<tr>
<th></th>
<th>-2 Log Likelihood</th>
<th>Chi-Square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>52.36</td>
<td>0.00</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>55.32</td>
<td>2.96</td>
<td>4</td>
<td>.56</td>
</tr>
<tr>
<td>RD</td>
<td>65.51</td>
<td>13.15</td>
<td>4</td>
<td>.01</td>
</tr>
<tr>
<td>Phonetic forms</td>
<td>247.92</td>
<td>195.56</td>
<td>6</td>
<td>.00</td>
</tr>
<tr>
<td>Semantic roles</td>
<td>92.75</td>
<td>40.39</td>
<td>14</td>
<td>.00</td>
</tr>
<tr>
<td>Syntactic roles</td>
<td>1036.38</td>
<td>984.02</td>
<td>10</td>
<td>.00</td>
</tr>
</tbody>
</table>

Nagelkerke = .98; R² = .97; MLOGIT not applicable

Table 0-7 *Accuracy of Syntactic Roles*

<table>
<thead>
<tr>
<th>Observed</th>
<th>Prediction</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Middle</td>
<td>Final</td>
<td>Accuracy</td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>101</td>
<td>0</td>
<td>0</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>0</td>
<td>435</td>
<td>53</td>
<td>89.1%</td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>0</td>
<td>0</td>
<td>387</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td>94.5%</td>
<td></td>
</tr>
</tbody>
</table>

Assuming that the accuracy equals the influential power of each predictor, i.e. topicality, phonetic forms, semantic roles and syntactic roles, upon the word order in Squiliq Atayal, we can get the radar chart in Figure 0-1, in which the word order of this language is viewed as the area inside the red quadrilateral. Obviously, this red quadrilateral is skewed to the point of syntactic roles, indicating that the most influential power comes from syntactic roles although the influential power of other components cannot be overlooked. If the influential power is further transformed into the scale of 100 percent, there appears the percentage of each component within the overall effect.
on word order, as shown in Figure 0-2. Syntactic roles have 34% of the overall effect, semantic roles 25%, phonetic forms 21%, and topicality 20%.

These two figures are meaningful to our linguistic interpretation in three aspects. First, they highlight again what we emphasize in the present thesis: all components/modules of grammar cooperate with each other in every single language. The study of the present thesis further visualizes their effect behind the word order of Squliq Atayal and manifests their effect into percentage. Second, they offer solid evidence for internal comparison. Linguists always get into the trouble about how to demonstrate the continuum of language facts. The most common strategy is adopting the one-dimensional line. One-dimensional lines only work well on the interplay of two internal variables, not on that of more than two variables. The following two-dimensional figures successfully demonstrate the interplay of more than two internal variables, and based on them, we are confident to say that syntactic roles are the most important factor behind word order because their solo power is up to 34% of the total effect. Last, these figures provide us the basis for typological or diachronic comparisons in the future. The pattern in Figure 0-1 shows the properties of word order in Squliq Atayal. What about discourse-configurational languages like Hungarian (É. Kiss 1987, 2002) and non-configurational languages like Warlpiri (Hale 1983, Nash 1986, Simpson 1991)? It is possible that the red quadrilateral of these languages is skewed to topicality to a bigger extent than it is in Squliq Atayal. With regard to diachronic comparisons, for instance, the word order of English gradually becomes rigid along the historical development (Allen 1995, Fischer & van der Wurff 2006, inter alia, Fischer et al. 2000), which definitely involves the change of the interplay of different components. It is possible that the skewedness to topicality diminishes in the history of the English language while the skewedness to syntactic roles increases.
A1.4 Summary

This appendix adopts the MLOGIT analysis to build up several models with various accuracy degrees of the word order prediction by one or several components such as syntactic roles, semantic roles, topicality (RD and TP), and phonetic forms. Only the full model containing all these predictor variables reaches the highest accuracy, which buttresses again the essentialness of the cooperation among different components in our language system. These models further show that the effect of syntactic roles is the greatest though the effect of others cannot be ignored. By visualizing the interplay
with radar charts and manifesting their effects into percentage, we create the basis for further typological and diachronic comparisons. However, the restrictions of sampling in our study may influence the generalization of our discoveries unless it is assumed that every native speaker of Squilq Atayal shares the same intuition.
References


Angouri, Jo. 2010. Quantitative, qualitative or both? Combining methods in linguistic


Carnie, Andrew. 2010. *Constituent Structure*: Oxford Surveys in Syntax and


Chang, Henry Yungli. 2003. AF verbs: transitive, intransitive, or both? In 2nd Workshop on Formosan Languages: On the Notion of "Verb" in Formosan Languages. Academia Sinica, Taipei.


In Honor of Dr. Li Fang-kuei on His 65 Birthday, 291-325. Taipei: Academia Sinica.


Northeastern Linguistic Society (Volume 11), 125-142.


Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge,
MA: The MIT Press.


382


Li, Paul Jen-kuei. 1985. Linguistic criteria for classifying Atayalic dialect groups.


Li, Thoi-yen. 2016. *The Literation of Taiwanese Aboriginal Languages*. Taipei: Chengchi University Press. [Chinese]


Newmeyer, Frederick J. 1998. Language Form and Language Function. Cambridge,
MA: The MIT Press.


Sapir, Edward. 1921. Language: An Introduction to the Study of Speech. New York:
Harcourt, Brace.
Starosta, Stanley. 1999. Transitivity, ergativity and the best analysis of Atayal case


Teng, Shou-hsin. 2012[1985]. Functional Grammar and Information Structure in


Tsai, Wei-Tien Dylan. 2015. A tale of two peripheries: evidence from Chinese adverbials, light verbs, applicatives and object fronting. In *The Cartography of


Wedgwood, Daniel, Gergely Pethő, and Ronnie Cann. 2006. Hungarian 'focus position' and English *i*-clefts: the semantic underspecification of 'focus' readings.

Wei, Jingyuen, Shuang He, and Yungxian Lo. 2011. *A Reference Grammar of Yanqi*
Zhuang. Beijing: China Social Sciences Press. [Chinese]


