

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

### **Evaluating competing models of codeswitching with reference to Mandarin/Tsou and Mandarin/Southern Min Data.**

Wang, Sung-Lan

*Award date:*  
2007

*Awarding institution:*  
University of Wales, Bangor

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

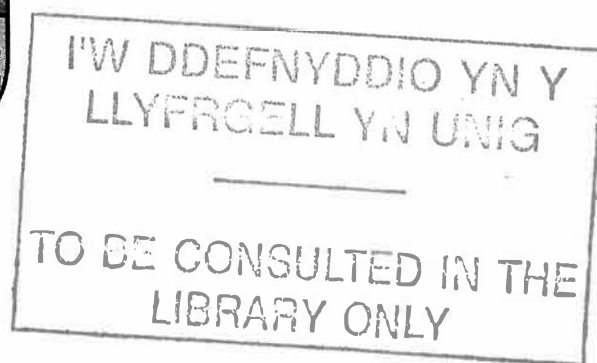
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Evaluating Competing Models of Codeswitching with Reference to  
Mandarin/Tsou and Mandarin/Southern Min Data**

• PRIFYSGOL CYMRU •  
UNIVERSITY OF WALES  
**BANGOR**



By

**Wang Sung Lan (Ted)**

**Supervisor**

**Professor Margaret Deuchar**

**A thesis submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in Bilingualism**

**University of Wales, Bangor**

**April, 2007**



【夫言非吹也，言者有言其所言者，特未定也。果有言邪？其未嘗有言邪？其以為異於鷦音？亦有辨乎？其無辨乎？道惡乎隱，而有真偽？言惡乎隱，而有是非？道惡乎往而不存，言惡乎存而不可？】

--- 莊子 齊物論

**The Adjustment of Controversies --- by Zhuang Zi (369 BC – 286BC)**

“But speech is not like the blowing (of the wind); the speaker has (a meaning in) his words. If, however, what he says, be indeterminate (as from a mind not made up), does he then really speak or not? He thinks that his words are different from the chirpings of fledgelings; but is there any distinction between them or not? But how can the Tâo be so obscured, that there should be 'a True' and 'a False' in it? How can speech be so obscured that there should be 'the Right' and 'the Wrong' about them? Where shall the Tâo go to that it will not be found? Where shall speech be found that it will be inappropriate? Tâo becomes obscured through the small comprehension (of the mind), and speech comes to be obscure through the vain-gloriousness (of the speaker). So it is that we have the contentions between the Literati and the Mohists, the one side affirming what the other denies, and vice versa. If we would decide on their several affirmations and denials, no plan is like bringing the (proper) light (of the mind) to bear on them”

(Translated by James Legge 1891; in *The Sacred Books of the East* Vols. 39 & 40, Oxford: Oxford University Press, 1891; reissued New York: Dover, 1962)

## **Abstract**

This thesis has two major aims. The first aim is to answer the research question: is there a universally applicable codeswitching model? The Matrix Language Frame (MLF) model proposed by Myers-Scotton (2002) has been successfully applied to different language pairs, and is claimed to be universally applicable. Muysken (2000), however, argues against the existence of any single codeswitching model with universal applicability, and proposes a typological approach to categorise the patterns of all codeswitching phenomenon. To fulfill the first aim, these two prominent codeswitching models were tested with the Mandarin/Tsou and Mandarin/Southern Min CS data, which were collected from informal conversations by bilingual Mandarin/Tsou and Mandarin/Southern Min speakers located in the Alisan and Tainan regions in Taiwan. The results revealed difficulties in applying both models to the data. To solve these problems, modified versions of both models were tested. The results of the re-analysis showed that the modified MLF model could be successfully applied to most of the Mandarin/Tsou and Mandarin/Southern Min data, while the modified typological model was still problematic. Since the original MLF model has been successfully applied to many other language pairs and its theoretical problem has been solved by my modification, I then argue that the revised MLF model proposed in this thesis is a universally applicable codeswitching model.

The second aim of this thesis is to predict the future of an endangered language, namely Tsou. To fulfill this aim, further qualitative and quantitative analyses of the results of the re-analysis of the Mandarin/Tsou data by using the revised MLF model were conducted. The Matrix Language (ML) Turnover Hypothesis proposed by Myers-Scotton (1998) was adopted to interpret the results of the analyses. The results showed that an ML turnover was complete and language shift was in progress in the Tsou language community. Hence, I argue that the survival of Tsou will be problematic.

## **Acknowledgement**

First, I have to express my gratitude to my supervisor Professor Margaret Deuchar and my second supervisor Dr. Pamela Macdonald. Without their guidance and encouragement, it would be impossible for me to finish this work. My thanks must go to all members of staff of the Linguistic department. I really appreciate their kindness and help throughout the period of my study in Bangor. I would also like to thank Mr. Eric Liu and Miss Cary Lo in the University of Newcastle upon Tyne for their encouragement and support. My special thanks must go to Professor Davis Anthony Scott who kindly read my thesis and gave me valuable comments. Finally, I would like to thank my parents, sister and all the relatives (especially to my grandmother Mrs. Li Zho Yu-lan who passed away in January, 2002) in Taiwan for their support.

## Table of Contents

<b>Acknowledgement.....</b>	<b>ii</b>
<b>Abstract.....</b>	<b>iii</b>
<b>Table of Contents.....</b>	<b>iv</b>
<b>Abbreviations and Gloss.....</b>	<b>xi</b>
<b>Figures and Tables.....</b>	<b>xiii</b>

### **Chapter 1: Introduction**

1.1. Introduction.....	1
1.2. Three Different Approaches to CS.....	2
1.2.1. A Sociolinguistic Approach.....	2
1.2.2. A CA (Conversational Analysis) Approach.....	2
1.2.3. A Structural Approach.....	3
1.3. Two Competing Models: The MLF Model vs. The Typological Approach.....	4
1.4. The Aims of This Thesis.....	5
1.5. Organisation of This Thesis.....	7

### **Chapter 2: Background**

2.1. Introduction.....	9
2.2. The Language Groups in Taiwan.....	11
2.2.1. Austronesian Aborigines.....	11
2.2.1.1. Overview.....	11
2.2.1.2. Austronesian Aborigines in Taiwan- Formosan Subgroups.....	13
2.2.1.3. Tsou.....	16
2.2.2. Southern Min and Hakka.....	17
2.2.3. Mainlanders.....	17
2.3. Three Waves of Immigration to Taiwan.....	18
2.3.1. The First Wave of Immigration: The Rule of General Zheng, Cheng-Kong (1662 – 1683) .....	18
2.3.2. The Second Wave of Immigration: The Rule of The Qing Dynasty (1683 – 1894) .....	18
2.3.3. The Third Wave of Immigration: The Defeat of The Nationalist Party (KMT –Kuomingtang) in the Chinese Civil War (1945 –1949).....	18
2.4. The Colonisation of Taiwan by Japan (1895 – 1945).....	19
2.4.1. Education and Language Policies --- Assimilation and Japanisation .....	19
2.4.2. The Overall Impacts of Colonisation by Japan on the Taiwanese People Today	

and the Current Linguistic Situation .....	22
2.5. The Rule of the KMT Party (1945 –2000) .....	24
2.5.1. A Brief Review.....	24
2.5.2. Language Policy.....	25
2.5.2.1. The Eradication of Japanisation and the Recovery of Chinese Connection (1945 –1969).....	26
2.5.3.2. The Continuation of the National Language Movement and the Promotion of Nationalism (1970 – 1986).....	28
2.5.3.3. Multi-cultural and Multilingual Policy (1987 –2000).....	26
2.6. The Rule of the DPP Party and Current Language Policy (2000-2007).....	30
2.7. Conclusion.....	31

### **Chapter 3: Review of the Literature**

3.1. Introduction.....	33
3.2. Various Approaches to CS .....	34
3.2.1. A Sociolinguistic Approach.....	34
3.2.1.1. Evaluation .....	36
3.2.2. A CA Approach.....	36
3.2.2.1. Evaluation .....	40
3.2.3. A Structural Approach to CS.....	41
3.2.3.1. Poplack’s Free Morpheme and Equivalence Constraints .....	42
3.2.3.1.1. Evaluation.....	45
3.2.3.2. The Government Constraint .....	47
3.2.3.3. A Minimalist Approach to CS .....	53
3.2.3.3.1. Evaluation.....	56
3.3. The MLF (Matrix Language Frame) Model.....	57
3.3.1. The Main Arguments of the MLF Model .....	57
3.3.1.1. Content-system Morpheme Opposition .....	57
3.3.1.2. Matrix Language-Embedded Language Opposition .....	58
3.3.2. The 4-M Model.....	60
3.4. The Typological Approach .....	63
3.4.1. Insertion.....	63
3.4.2. Alternation .....	67
3.4.3. Congruent Lexicalization .....	68
3.4.4. CS Patterns and Extra-linguistic Factors.....	70
3.5. A Cross-disciplinary Approach .....	71
3.6. Discussion of Two Morphosyntactic Approaches.....	75
3.6.1. A Contrastive Analysis of the Two Approaches .....	76

3.6.2. Evaluation of the Two Approaches .....	77
3.6.2.1. The MLF Model.....	77
3.6.2.2. The Typology Approach .....	81
3.7. Earlier CS Research Relevant to Chinese.....	82
3.7.1. An Integrated Approach to Language Choice Patterns and Chinese/English CS .....	82
3.7.2. Tonal Aspects to Chinese-English CS .....	84
3.7.3. Morphosyntactic Approach to Chinese-English CS .....	90
3.8. Conclusion .....	93

#### **Chapter 4: An Introduction of Mandarin Chinese, Southern Min, and Tsou**

4.1. Introduction .....	94
4.1.1. Aims of this Chapter.....	96
4.1.2. The Organisation of This Chapter .....	98
4.2. Word Order .....	99
4.2.1. Mandarin.....	99
4.2.2. Southern Min.....	104
4.3. Aspect.....	107
4.3.1. The Perfective Aspect .....	107
4.3.1.1. Mandarin Perfective Aspect Marker: “le”.....	107
4.3.1.2. Southern Min Perfective Aspect Marker: “liao” and “a”.....	109
4.3.2. The Durative Aspect .....	110
4.3.2.1. Mandarin Durative Aspect Markers: “zai” and “zhe” .....	110
4.3.2.2. Southern Min Durative Aspect Markers: “di”.....	111
4.3.3. The Experiential Aspect .....	112
4.3.3.1. Mandarin Experiential Aspect Marker: “guo”.....	112
4.3.3.2. Southern Min Experiential Aspect Marker: “gue”.....	112
4.3.4. The Delimitative Aspect .....	113
4.3.4.1. Mandarin Delimitative Aspect.....	113
4.3.4.2. Southern Min Delimitative Aspect: “lei”.....	113
4.4. Classifiers.....	114
4.5. Possessive Marker and Nominalizer.....	120
4.5.1. Mandarin .....	120
4.5.1.1. Possessive Marker .....	120
4.5.1.2. Nominalizer .....	121
4.5.2. Southern Min.....	122
4.5.2.1. Possessive Marker .....	122
4.5.2.2. Nominalizer .....	123



4.6. Serial Verbs.....	124
4.7. Sentence Final Particles .....	127
4.7.1. The Particle “a” .....	128
4.7.1.1. Mandarin .....	128
4.7.1.2. Southern Min.....	130
4.7.2. The Question Particle “la”.....	133
4.7.2.1. Mandarin .....	133
4.7.2.2. Southern Min .....	134
4.7.3. The Question Particle ‘ma’ .....	135
4.7.3.1. Mandarin.....	135
4.7.3.2. Southern Min .....	135
4.8. The <i>A-not-A</i> Question vs. <i>A...not</i> Question Construction.....	136
4.8.1. The <i>A-not-A</i> Question in Mandarin.....	136
4.8.2. The <i>A...not</i> Question in Southern Min.....	137
4.9. A Brief Description of Tsou Grammar .....	138
4.9.1. The Case Marking System.....	139
4.9.1.1. [+/- vis]: “?e” vs. “?o” and “da” vs. “do”.....	142
4.9.2. Word Order .....	143
4.9.3. Tense and Thematic Role Marking System.....	144
4.9.3.1. Non-Past Tense and Agent Markers: “mi-” and “mio” .....	146
4.9.3.2. Past Tense and Agent Markers: “mo/moh-” and “moso”.....	147
4.9.3.3. Non-agent Tense Marker: “o/oh-” and “i-” .....	148
4.9.4. Summary .....	149
<b>Chapter 5: Data Collection</b>	
5.1. Aims.....	151
5.2. Educational Level and Age as Controlled Variables .....	152
5.3. Organisation .....	153
5.4. Mandarin/Tsou Community.....	154
5.4.1. Geographical Distribution of Tsou Villages .....	154
5.4.2. Linguistic Situation.....	156
5.4.3. Group 1: Older Tsou People (age 55 – 70).....	157
5.4.3.1. Location: Tapangu.....	157
5.4.3.2. Subjects and Settings .....	158
5.4.3.3. Data Recording.....	160
5.4.3.4. Data .....	161
5.4.4. Group 2: Younger Tsou People (age 25 – 40).....	162
5.4.4.1. Location- Cayamavana Village.....	164

5.4.4.2. Subjects .....	164
5.4.4.3. Data Recording.....	165
5.4.4.4. Data .....	166
5.5. The Mandarin/Southern Min Community.....	166
5.5.1. Tainan Area.....	166
5.5.2. Group 3: Educated Below University Level/Older People (age50 –60).....	169
5.5.2.1. Subjects and Settings .....	169
5.5.2.2. Meeting The Subjects .....	170
5.5.2.3. Data Recording .....	171
5.5.2.4. Transcribing Data.....	171
5.5.2.5. Transcribing Procedures .....	172
5.5.3. Group 4: University Level/Older People (age 50 –60).....	172
5.5.3.1. Subjects and Settings.....	172
5.5.3.2. Meeting The Subjects.....	173
5.5.3.3. Data Recording .....	174
5.5.3.4. Transcribing Data.....	174
5.5.4. Group 5: University Student (age 18 –25).....	174
5.5.4.1. Subjects .....	174
5.5.4.2. Meeting the Subjects.....	175
5.5.4.3. Data Recording.....	176
5.5.4.4. Transcribing Data .....	176

## **Chapter 6: Data Analysis**

6.1. Introduction .....	177
6.2. Testing The MLF Model .....	179
6.2.1. Mandarin/Tsou Data .....	180
6.2.1.1. Method of Analysis.....	180
6.2.1.2. Results .....	183
6.2.2. Mandarin/Southern Min Data .....	184
6.2.2.1. Theoretical Problems of The MLF Model.....	185
6.2.2.2. Aspect Markers as Outsider Late System Morphemes? .....	190
6.2.2.3. Other Possible Criteria .....	192
6.2.3. A Revised Version of The MLF Model .....	202
6.2.3.1. Stage 1: The Morpheme Order Principle and the System Morpheme Principle.....	204
6.2.3.2. Stage 2: The Morpheme Counting Principle.....	204
6.2.3.3. Stage 3: The Uniform Structure Principle.....	212
6.2.3.4. The Application of the Revised MLF Model.....	212

6.3. Testing The Typology Approach.....	219
6.3.1. Review of The Typology Approach.....	219
6.3.2. Theoretical Problem --- How to define a switch?.....	222
6.3.3. A Revised Typological Approach .....	224
6.3.3.1. CP as A Basic Unit of Analysis.....	225
6.3.3.2. The Application of Muysken’s (2000) Features.....	227
6.4. Summary.....	238

## **Chapter 7: Re-analysis of the Data Using the Revised MLF Model**

7.1. Introduction .....	241
7.2. An Evaluation of The Revised MLF Model.....	244
7.2.1. Mandarin/Tsou Data .....	244
7.2.2. Mandarin/Southern Min Data.....	245
7.2.3. Mandarin/Tsou vs. Mandarin/Southern Min Data: A Comparison of the Results.....	248
7.3. Practical Implications of The Results of Analysis .....	249
7.3.1. The Matrix Language (ML) Turnover Hypothesis .....	249
7.3.1.1. Main Arguments .....	249
7.3.1.2. An ML Turnover in Mandarin/Tsou CS.....	256
7.3.2. EL (Embedded Language) Islands and EL Proficiency.....	260

## **Chapter 8: Summary & Conclusion**

8.1. Summary and Conclusion.....	272
8.2. Concluding Remarks.....	273
8.2.1. The Universal Applicability of the Revised MLF Model.....	273
8.2.2. The Future of the Endangered Language Tsou.....	275
8.3. Limitations and Suggestions for Future Studies.....	276

<b>References.....</b>	<b>280</b>
<b>Appendix 1: Letter of Consent.....</b>	<b>286</b>
<b>Appendix 2: Letter of Consent (Chinese Version).....</b>	<b>287</b>
<b>Appendix 3: Questionnaire.....</b>	<b>288</b>
<b>Appendix 4: Questionnaire (Chinese version).....</b>	<b>290</b>
<b>Appendix 5: Bilingual Mandarin/Tsou Clauses Collected from Group1 .....</b>	<b>292</b>
<b>Appendix 6: Bilingual Mandarin/Tsou clauses collected from Group 2.....</b>	<b>307</b>
<b>Appendix 7: Bilingual Mandarin/Southern Min clauses collected     from Group 3.....</b>	<b>312</b>

<b>Appendix 8:</b> Bilingual Mandarin/Southern Min clauses collected from Group 4.....	329
<b>Appendix 9:</b> Bilingual Mandarin/Southern Min clauses collected from Group 5.....	347

## Abbreviations and Gloss

1s = 1<sup>st</sup> person singular  
2s = 2<sup>nd</sup> person singular  
3s = 3<sup>rd</sup> person singular  
3p = 3<sup>rd</sup> person plural  
affirm = affirmation  
Agent = agent marker  
art = article  
asp = aspect marker  
class = classifier  
cla = clarification  
compl = complaint  
comt = comment  
cop = copular verb  
delim = delimitative aspect marker  
DM = discourse marker  
dur = durative aspect marker  
EQM = exclamation question marker  
excl = exclamation marker  
exp = experiential aspect marker  
expl = explanation marker  
gen = genitive case marker  
hab = habitual aspect marker  
inform = informative  
imp = imperfect tense  
impat = marking speaker's impatience  
Interj = interjection  
neg = negation marker  
Nom = nominative case marker  
nom = nominalizer  
Non-Agent = Non-agent marker  
NP = noun phrase  
Obl = oblique case marker  
part. = particle  
pass = past tense  
perf = perfective aspect  
plur = plural marker

PM =pronoun marker  
poss = possessive marker  
PP = Prepositional Phrase  
Ques. = question marker  
rep =report  
req = request  
RF = reducing forcefulness  
RM = relative clause marker  
sug = suggestion  
surp = surprised  
SVO = Subject Verb Object  
Unexpected = an unexpected event  
unfor = unfortunate event  
VOS = Verb Object Subject  
VP = Verb phrase

## Figures

Figure 1: Taiwan Strait Area.....	10
Figure 2: The Austronesian Family.....	12
Figure 3: The Uppermost Nodes of The Austronesian Genealogical Tree.....	13
Figure 4: The Twelve Austronesian Aboriginal Tribes and Their Geographical Distribution in Taiwan.....	15
Figure 5: Current Tsou Subgroups and Dialects.....	16
Figure 6: The Map of Taiwan.....	155
Figure 7: The Geographical Distribution of Tsou Villages.....	156
Figure 8: Taiwan Strait Area.....	168

## Tables

Table 1: Mandarin and Southern Min Classifiers with Glosses in English.....	115
Table 2: The Semantic Features of Tsou Case Markers.....	141
Table 3: The Tense and Thematic Role Marking System in Tsou.....	145
Table 4: Results of The Application of MLF Model to Mandarin/Tsou Data.....	184
Table 5: Results of The Application of MLF Model to Mandarin/Southern Min Data.....	188
Table 6: Mandarin and Southern Min Classifiers with Glosses in English.....	194
Table 7: A Revised Version of The MLF Model.....	203
Table 8: Results of Applying the Revised MLF Model to Mandarin/Southern Min Data.....	217
Table 9: Muysken's View of the Relation between Code-switching Patterns and Extra-linguistic Factors.....	222
Table 10: A Revised Version of The MLF Model.....	225

Table 11: Diagnostic Features of the Three Patterns of CS.....	227
Table 12: The Quantitative Analysis of Data Using Muysken’s Features.....	230
Table 13: The Overall Results of Mandarin/Tsou and Mandarin/Southern Min Data.....	234
Table 14: Muysken’s View of the Relation between Code-switching Patterns and Extralinguistic Factors.....	236
Table 15: A Revised Version of The MLF Model.....	241
Table 16: Results of The Application of the Revised MLF Model to Mandarin/Tsou Data.....	245
Table 17: The Results of Applying the Revised MLF Model to Mandarin/Southern Min Data.....	246
Table 18: The Number and Percentage of EL Islands vs. Singly-occurring Lexemes in Mandarin/Tsou CPs with Mandarin as ML.....	264
Table 19: Nature of EL Elements, Group 1 vs. Group 2: EL islands vs. Singly- occurring Lexemes in Bilingual CPs with Mandarin as the ML.....	265
Table 20: The Number and Percentage of EL Islands vs. Singly-occurring Lexemes in Mandarin/Tsou CPs with Tsou as ML.....	267
Table 21: Nature of EL elements, Group 1 vs. Group 2: EL islands vs. Singly- occurring Lexemes in Bilingual CPs with Tsou as the ML.....	268
Table 22: A Revised Version of The MLF Model.....	272



## Chapter 1: Introduction

### 1.1. Introduction

Codeswitching (abbreviated as CS below) behaviour is one of the most common results of language contact, and has been studied from different perspectives by scholars from various disciplines. Two frequently asked research questions in the study of CS are: (1) why do people switch between different codes? (2) How do people code-switch? Linguists who want to answer the *why* question often investigate the relationship between code-switched utterances and certain extra-linguistic factors, for they argue that those extra-linguistic factors are the main causes of CS. For instance, a sociolinguistic approach is to study CS utterances with reference to certain sociolinguistic variables (e.g., age, gender, social class etc.). Practitioners of the CA (Conversational Analysis) approach investigate the conversational function or social meanings behind CS.

Researchers who intend to answer the *how* question often adopt a structural approach to examine the grammatical structure of code-switched utterances, and propose different grammatical rules or constraints on CS. Such an approach is adopted in this thesis. To my knowledge, the CS phenomena in the two language pairs in this thesis (Mandarin/Tsou and Mandarin/Southern Min<sup>1</sup>) are little studied. Hence, I argue that it is necessary to study the basic grammar or linguistic features of code-switched utterances before one can link CS to any extra-linguistic factors. Before discussing the

---

<sup>1</sup> "Southern Min" is also known as Taiwanese or Hokkien, and is a variety of Chinese spoken in the Fujian province (also known as Min) in Southern China. Since 76.2% of the entire population of Taiwan are ethnically southern Min, an ethnic group which was originally from the southern part of Fujian province, the language they speak (i.e., Southern Min) is often called "Taiwanese" today. However, 'Taiwanese' is also used in a broader sense to refer to all the ethnic groups (e.g., Austronesian aborigines, Hakka, Southern Min, and Mainlanders) in Taiwan. To avoid any confusion between these two terms, "Southern Min" is used in this thesis to refer to the language used by the ethnically southern Min group in Taiwan.

research aims of this thesis, we shall briefly review three dominant approaches to CS (i.e., sociolinguistic, CA, and structural) in the current literature.

## 1.2. Three Different Approaches to CS

### 1.2.1. A Sociolinguistic Approach

A sociolinguistic approach, for instance, often looks at CS phenomena with reference to different sociolinguistic variables (e.g., age, gender, social class etc.), and examines the social meanings behind it. In their study of CS in Hennesberget, a city in Northern Norway, Blom, & Gumperz (1972) argue that the local inhabitants switch between the standard language Bokmål and the dialect Ranamål depending on the social situation and the topics that they discuss. The standard is often used when speaking to non-local people (such as tourists) or discussing education or national affairs. The dialect, however is used when talking to local people or discussing private or local affairs. They go on to argue that the social meaning behind the use of the dialect is that it expresses local identity and group membership of the local community while the use of the standard denotes “officiality, expertise and politeness toward strangers (p.135)”. Blom & Gumperz’s study certainly provides valuable explanations as to why people switch between different codes.

### 1.2.2. A Conversational Analysis (CA) Approach

Auer (1998) argues that many macro-sociolinguistic investigations of CS restrict themselves to analysing the social meaning of CS by reference to its participants, topic or settings, but they fail to consider the structure of a conversation. Instead of a macro-sociolinguistic approach to CS, linguists such as Auer (2005), Li Wei (2005) and Gafaranga (2005) adopt a CA (Conversational Analysis) approach to investigate the relationship between CS and conversational structure. A turn-by-turn analysis of a

conversation is conducted by CA practitioners in order to answer questions such as: why is CS more likely to occur at certain sequential positions in a conversation? Does the occurrence of CS in conversational interaction have any particular communicative function? Hence, Li Wei (2005, p.382) argues that to those who adopt a CA approach to CS emphasise “how such things as identity, attitude and relationship are presented, understood, accepted, rejected or changed in the process of interaction”. He continues to argue that a CA approach to CS not only focuses on the analysis of conversational structure itself, but also attempts to link it to a wider extra-linguistic context. A CA approach also provides valuable insights into CS phenomena in terms of filling the gaps that the macro-sociolinguistic approach does not cover.

### 1.2.3. A Structural Approach

As mentioned above, the structural approach examines the structure of the code-switched utterances. Structural linguists such as Poplack (1980) and Myers-Scotton (1993/1997; 2002a) argue that CS utterances are rule-governed. For instance, Poplack (1980, p.615) argues that “CS is itself a discrete mode of speaking, possibly emanating from a single CS grammar composed of the overlapping sectors of the grammars of L1 and L2.” However, this raises a fundamental question: Are there universally applicable grammatical rules or constraints on CS? To answer this question, many linguists (e.g., Poplack 1980; DiSciullo, Muysken, & Singh 1986) proposed different constraints or grammatical theories and attempted to provide a universally applicable generalisation to CS phenomena (c.f. Chapter 3). However, these grammatical theories or constraints were soon rejected by other linguists with reference to their empirical CS data from different language pairs.

Myers-Scotton’s (2002a) Matrix Language Frame (MLF) model and Muysken’s (2000)

typological approach are the two most influential structural models in the current study of CS. These two models hold completely different views on the research question: are there any universally applicable CS constraints? The following section will briefly introduce the major arguments of these two models, and discuss what perspectives these two scholars have on answers to this question.

### 1.3. Two Competing Models: The MLF Model vs. the Typological Approach

The Matrix Language Frame (abbreviated to MLF below) model was first proposed by Myers-Scotton (1991), and was revised in Myers-Scotton (1993/1997; 2002a). It is one of the most prominent models in this field, and enjoys considerable popularity among linguists who are interested in a structural approach to CS. Myers-Scotton (2006, p.248) argues that “the MLF model is generally supported across various sets of CS data reported in other studies, with only a few apparent exceptions... We stress the universality of support, no matter which languages are involved”. The MLF model investigates the interaction between the grammatical systems (focusing on word order and inflectional morphology) of the participating languages within a bilingual CP, which is the basic unit of analysis. Myers-Scotton (2002a, p.54) defines a CP as “the projection of complementizer”. A bilingual CP is defined as “a CP which contains bilingual constituents” (p.56). The key assumption of this model is that the roles of the participating languages in a bilingual CP are not equal. In other words, the matrix language, which provides the basic morphosyntactic structure, makes a larger contribution while the other participating language (i.e., the embedded language) contributes less.

The typological approach proposed by Muysken (2000) is also an influential morphosyntactic model, but it looks at CS from a different perspective. In contrast to

Myers-Scotton's (2002a) point of view, Muysken (2000, p.3) argues that "I do not propose a single model of CS, since I do not think there is such a model..." Hence, instead of proposing a single grammatical model, he categorises all CS data according to three patterns i.e., **insertion**, **alternation** and **congruent lexicalization**, and each of these three is argued to be linked to various extra-linguistic factors (c.f. section 3.4.4). Unlike the MLF model which merely focuses on intra-clausal switches, Muysken's typological approach not only looks at intra-clausal CS, but also inter-clausal CS and the switches which occur between different turns. Hence, his model takes a broader perspective on the grammar of CS utterances.

#### 1.4. The Aims of this Thesis

In the discussion above, we showed that there is still no definite answer to the question: are there universal constraints on CS? One of the major aims of this thesis is to find a possible answer to this question. The MLF model, which particularly looks at the word order and inflectional morphology of CS utterances, is claimed to be universally applicable, and has been successfully applied to data from various language pairs e.g., Swahili/English (Myers-Scotton 1993/1997); German/English (Fuller 2000); Mandarin/English (Wei 2001); Spanish/English (Jake, Myers-Scotton, & Gross 2002); and Welsh/English (Deuchar 2006). What most of these language pairs have in common is that they are all typologically different languages (in terms of their word order) and arguably have a comparatively complex system of inflectional morphology (if compared to an isolating language such as Chinese), which is essential for the application of the MLF model (c.f. section 3.3.1.2). To my knowledge, the MLF model has not yet been tested with two typologically very similar language pairs (e.g., my Mandarin/Southern Min data) which share most of their syntax (both are SVO languages) but differ in phonology, and in which there is almost no inflectional

morphology. The linguistic nature of Mandarin and Southern Min may therefore pose some interesting challenges to the MLF model.

The other language pair (i.e., Mandarin/Tsou) in this study includes an endangered minority language, namely Tsou. Tsou is an Austronesian language spoken by one of the aboriginal tribes in Taiwan, and is linguistically very different from Mandarin. It has a VOS word order and abundant inflectional morphology. The study of Tsou is still at an initial stage, so little research has been conducted. Previous literature related to Tsou (e.g., Zeitoun 1992; Cho 1997; Chang 1998; Chang 2004; Huang 2005), all focus on investigating syntax, phonology, or semantics, and none of them is related to CS. By testing the MLF model with a language pair which has not yet been tested (i.e., the Mandarin/Tsou data) and with a language pair which has the same word order and very limited amount of inflectional morphology (i.e., Mandarin/Southern Min), this thesis aims to show whether or not this model is really universally applicable.

As mentioned earlier, Muysken's (2000) typological approach provides a broader view to CS phenomena if compared with the MLF model. Moreover, it holds an opposite point of view to the MLF model on the issue of the universality of CS constraints. Hence, I plan to test his model and compare the results with those of the MLF model.

In addition to the theoretical motivation for testing the two important CS models, I have a more practical reason to conduct this research, namely to predict the future of an endangered language (i.e., Tsou). Myers-Scotton (1998; 2002a) argues that one of the hypotheses associated with the MLF model, namely the Matrix Language Turnover Hypothesis (c.f. section 7.3.1.), could be used to predict the death of an

endangered language. Matrix language turnover refers to the situation in which the original matrix language (i.e., the language which provides the morphosyntactic structure) of a bilingual CP becomes the embedded language and vice versa (Myser-Scotton 1998). In many cases, the original matrix language is often the endangered language while the original embedded language is the language with higher prestige. Because of the influence of various socio-political factors (e.g., strict monolingual language policy), the more prestigious language may replace the endangered language (either gradually or rapidly) as the matrix language of most of the CS utterances produced by the speakers of that endangered language. Myers-Scotton argues that if such a process is completed, it is very likely that language shift will follow, and therefore the survival of that endangered language will be problematic. Since Tsou is in an endangered situation, it is hoped that the present study can gain some insights into its future by testing the MLF model and applying the Matrix Language Turnover Hypothesis.

### 1.5. Organisation of this Thesis

The overall organisation of this thesis is given in this section. In Chapter 2, I will provide some basic historical, political, and geographical information about Taiwan, and briefly describe the various ethnic groups in the island and the languages they speak.

In Chapter 3, I will briefly review different approaches to CS in the current literature. The focus will be on a detailed introduction of these two morphosyntactic models (i.e., the MLF model and the typological model) that this thesis will test. In Chapter 4, I will discuss the syntactic and morphological systems of the three languages, namely Mandarin, Southern Min, and Tsou. In Chapter 5, how my fieldwork was conducted

will be described. It will discuss how the sample groups were composed and include detailed information about each group. Moreover, the detailed data collection procedures will also be given.

Chapter 6 is the core of this thesis and will test these two morphosyntactic models with my Mandarin/Tsou and Mandarin/Southern Min data. Detailed information as to how the analysis was conducted will be given. A discussion of the results obtained from testing the two CS models, namely Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach, will be included. Then, some theoretical problems of these two models will be discussed. Finally, I will try to modify the original MLF model based on the nature of my data, and propose a revised model.

In Chapter 7, I will first evaluate the revised MLF model by re-analysing the Mandarin/Tsou and Mandarin/Southern Min data. Then, sociolinguistic interpretations and practical application of the results of the re-analysis will be discussed. Finally, a summary and the conclusions of this thesis will be provided in Chapter 8. This will be followed by some discussion of the limitations of this thesis as well as the recommendations for future studies.



## Chapter 2: Background

### 2.1 Introduction

Taiwan, also known as Formosa, is an island with an area of about 36,000 km<sup>2</sup> in East Asia with Japan to the north and the Philippines to the south. It has been a colony of Holland (1624–1662), of Spain (1626–1642) and of Japan (1895–1945) (Huang 2000). It is geographically separated from Mainland China by the Taiwan Strait which is approximately 100 miles wide. Politically, it is ruled by a democratic regime, known as the ROC (Republic of China), and is distinct from communist China, which is known as the PRC (People’s Republic of China) on Mainland China. Taiwan claims to be an independent country, but this is not internationally recognised. Despite its political separation from Mainland China, the culture, social customs, and languages spoken in Taiwan (apart from the Austronesian languages), have strong links to those of the Mainland. The official language and the national language of Taiwan is the same as that of Mainland China, namely Mandarin Chinese.

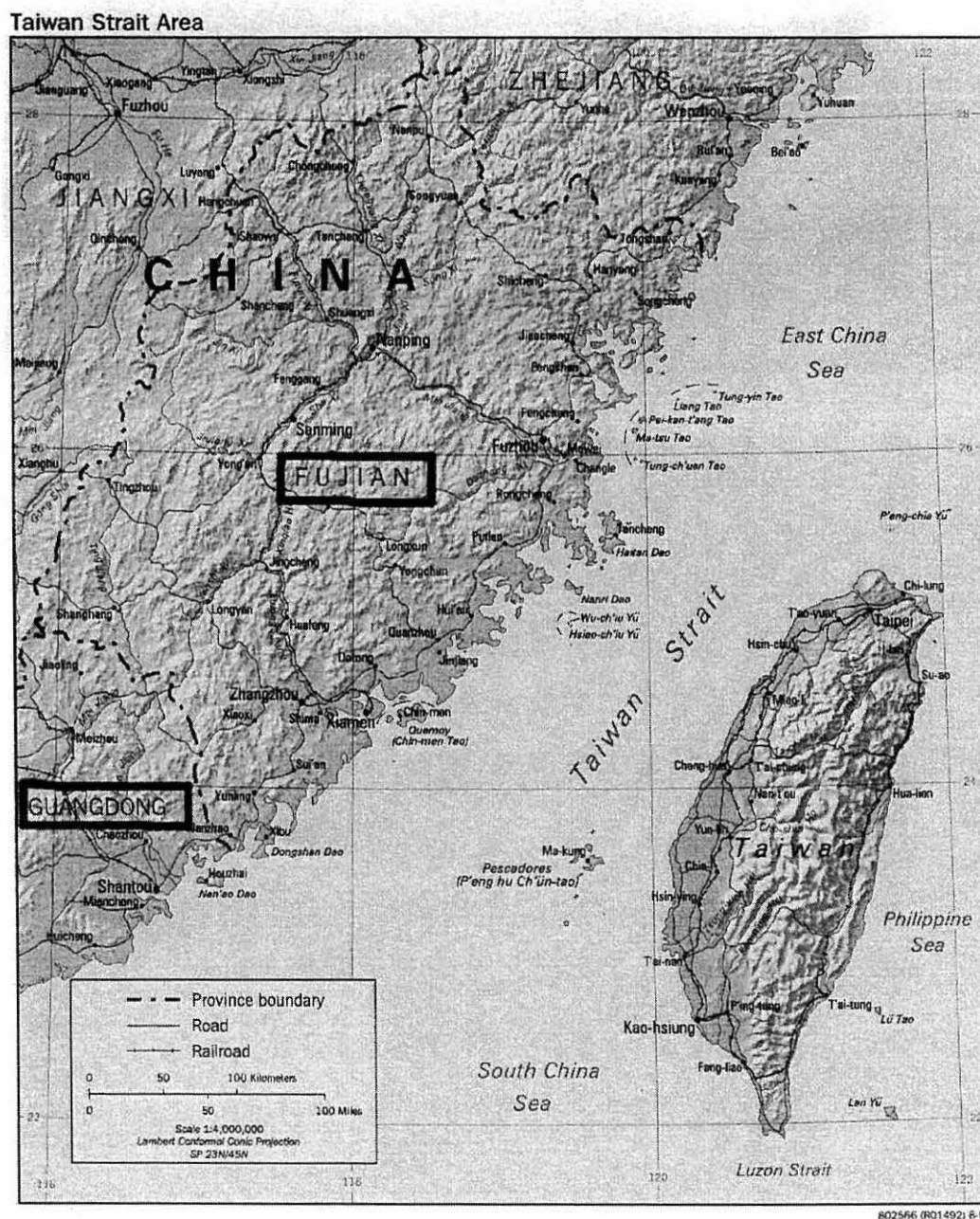
The total population of Taiwan is around 23 million, which includes a variety of groups, i.e., Austronesian aborigines, Southern Min, Hakka and the so-called “mainlanders<sup>2</sup>”. The languages spoken by these groups differ. Archaeological evidence suggests that the Austronesian aborigines were probably in Taiwan by the late Palaeolithic period. These Austronesian aborigines can be further divided into twelve main tribes, and each of them has a distinct language. Southern Min, Hakka and mainlanders are settlers from Mainland China who arrived at different points in time. They speak several varieties of Chinese, including Southern Min (also known as Taiwanese or Hokkien), Hakka, Xichuanese etc. The speakers of these varieties are

---

<sup>2</sup> “Mainlanders” refers to the group of people (mainly soldiers and their relatives from different provinces in China) who fled to Taiwan with the Nationalist Party when it lost the Chinese Civil War (1945–1949).

not mutually intelligible; hence, Mandarin is used as the lingua franca. The colonisation by Japan (1895–1945) brought the Japanese language into Taiwan, and increased the complexity of the linguistic situation.

**Figure 1: The Taiwan Strait Area** (Map Source: The University of Texas, Austin: [http://www.lib.utexas.edu/maps/cia04/taiwan\\_sm04.gif](http://www.lib.utexas.edu/maps/cia04/taiwan_sm04.gif); Access Date: April 10, 2005)



In this chapter, I shall first provide background information on these four groups, e.g.,

the origin and geographical distribution of the Austronesian aborigines, and the origins of the other three groups. Since Southern Min and Tsou are the focuses of this study, a more detailed description of these two groups will be given. Second, since Taiwan was ruled by different colonial powers and Chinese regimes, I shall briefly describe them in a chronological order. Third, a discussion of the language policies implemented by the Japanese government and the succeeding Chinese regimes will be given.

## 2.2. The Language Groups in Taiwan

### 2.2.1. Austronesian Aborigines

#### 2.2.1.1. Overview

The geographical distribution of the Austronesian language family is very wide, starting from the island of Madagascar on the east coast of Africa to Easter Island and extending to Taiwan, Northern Australia and New Zealand, and most of the Melanesian and Polynesian islands in the Pacific Ocean (Ruhlen 1987). This is illustrated in Figure 2 below.

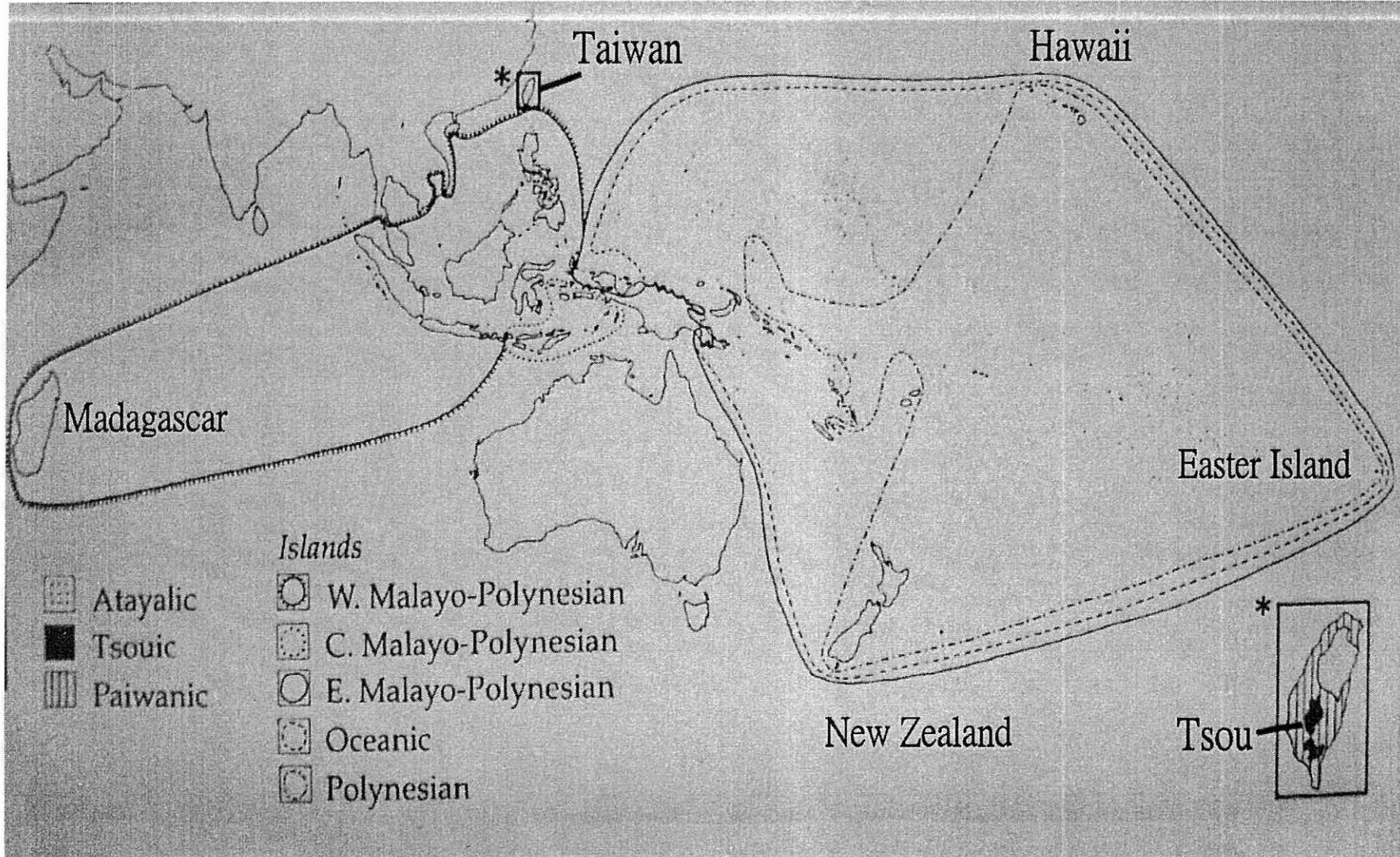
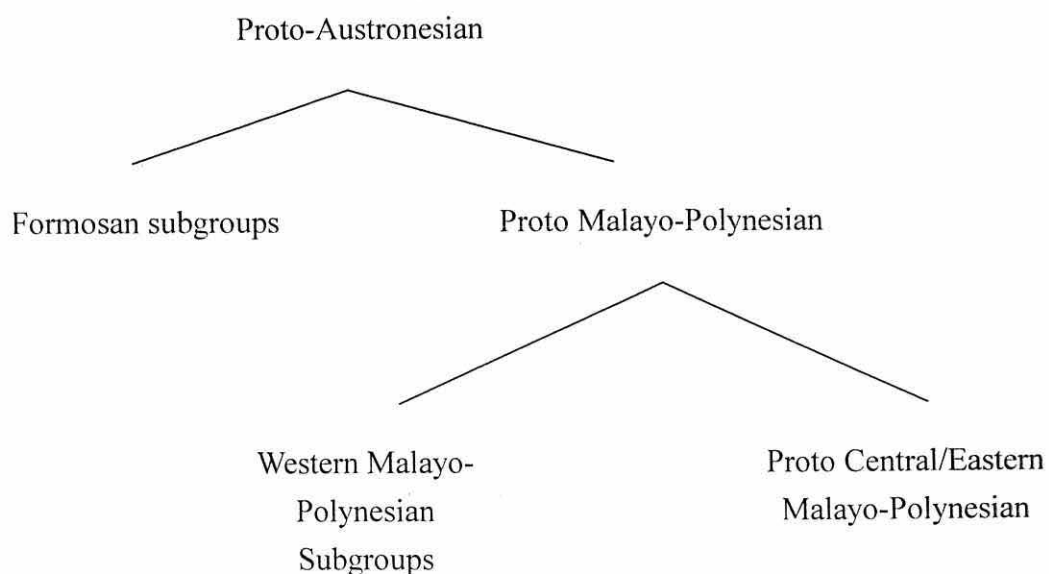


Figure 2: The Austronesian Family (Ruhlen 1987, p. 160)

The current archaeological, biological and comparative linguistic evidence places the possible origin of this language family in a number of regions: (1) the southeast coast of the Asian continent (Haudricourt 1954; cited in Li 1999) (2) the eastern part of New Guinea (Dyen 1965) (3) Taiwan (Blust 1985; Bellwood 1991). Although the exact of origin of this language family is still not clear, many researchers (Haudricourt 1954, cited in Chang 2004; Tsuchida 1976; Blust 1985; Li 1999) agree that the Austronesian languages in Taiwan (i.e., the Formosan subgroups) are probably the oldest in this language family. This is because the Formosan languages are at the uppermost position in an Austronesian language family tree (c.f. Figure 3), and many of their phonemic features correspond to those of the Proto-Austronesian language compared to other Austronesian languages.

**Figure 3: The Uppermost Nodes of the Austronesian Genealogical Tree (Chang 2004, p.3)**



#### 2.2.1.2. Austronesian Aborigines in Taiwan - Formosan Subgroups

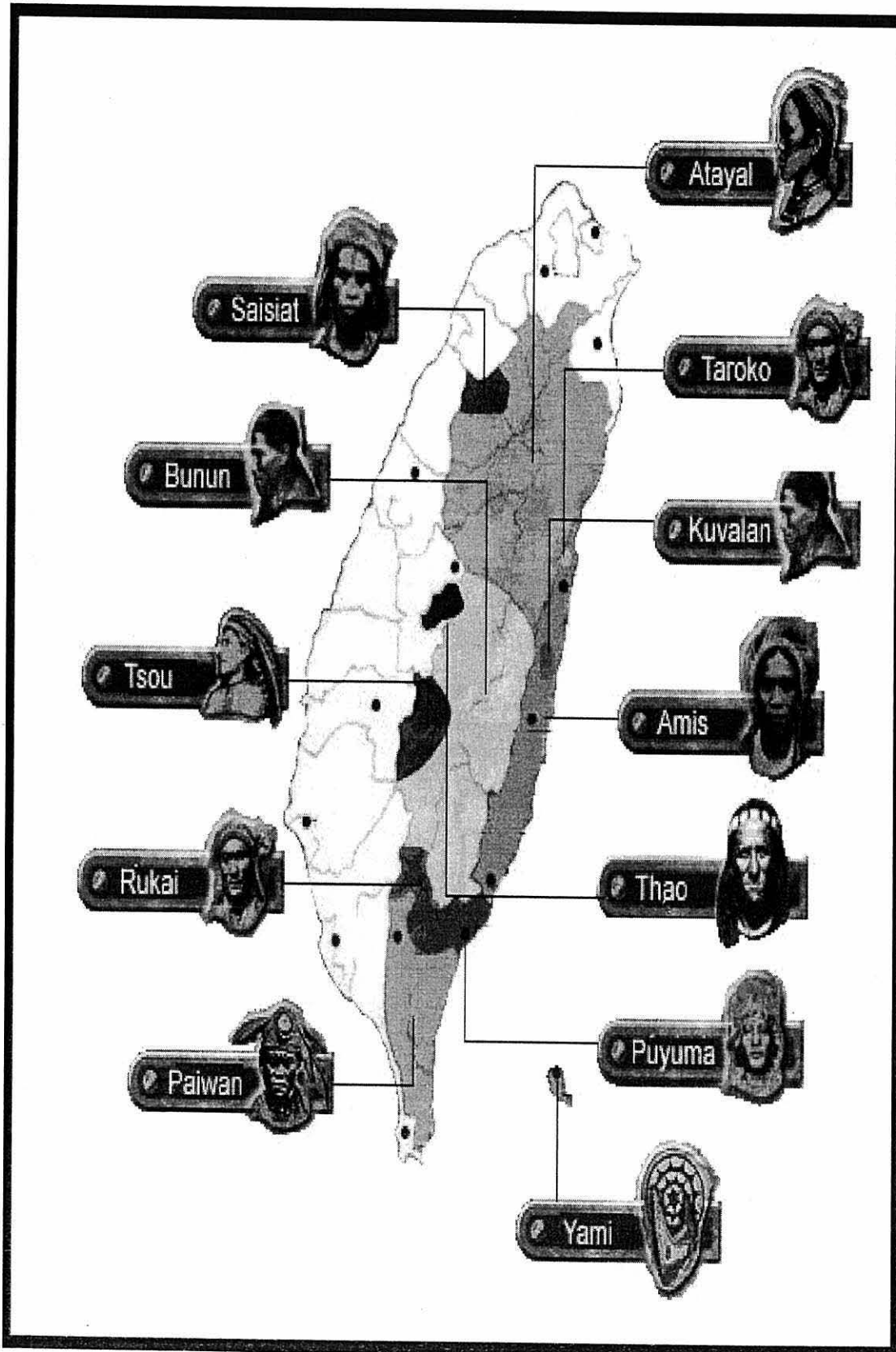
According to archaeological studies, the earliest human activities in Taiwan were in

the late Palaeolithic period. Although there is still no direct evidence linking these activities to the Austronesians, it is generally believed that these prehistoric human beings may have been the ancestors of the aborigines in Taiwan or even all the Austronesian aborigines (Li 1999).

About 1.8% of the population in Taiwan are Austronesian aborigines made up of fourteen different tribes, but only twelve of them are well-known (Chun 2005). These tribes are: Saisiat, Bunun, Tsou, Rukai, Paiwan, Atayal, Taroko, Kuvalan, Amis, Thao, Puyuma and Yami. The languages these twelve tribes speak can be categorised into three main groups: Atayalic, Tsouic, and Paiwanic (Ruhlen 1987). Their geographical distribution is shown in Figure 4. In this figure, the marked areas are mainly mountainous regions where these aborigines live. The unmarked areas are mainly plains where other aboriginal tribes used to live; these tribes are now extinct after long-term contact and intermarriage with Chinese immigrants (Li 1999).

**Figure 4: The Twelve Austronesian Aboriginal Tribes and Their Geographical Distribution in Taiwan**

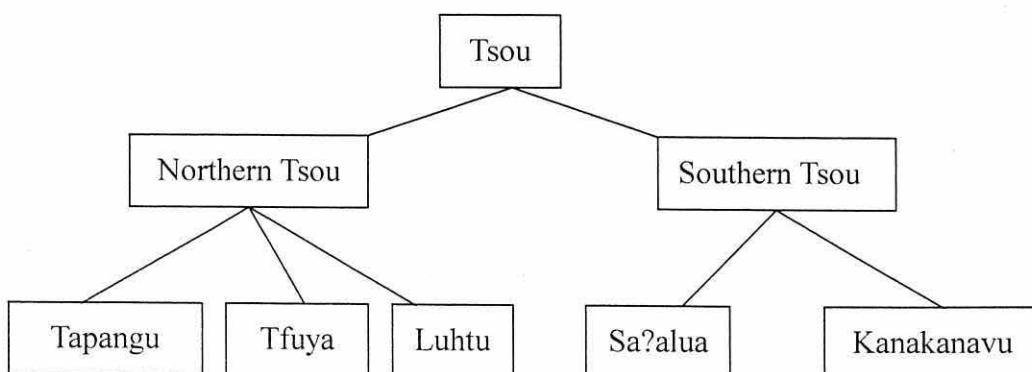
(Map Source: Council of The Indigenous People, Taiwan: [www.taiwan.gov.tw/INTRO/FMINTRO](http://www.taiwan.gov.tw/INTRO/FMINTRO))



### 2.2.1.3. Tsou

The current population of Tsou is about 6,000 (Chang 2004). They live in the mountainous areas in central Taiwan at heights between 400 and 1750 metres. It is generally believed that the Tsou tribe can be further divided into two sub-groups: Northern Tsou and Southern Tsou (Li 1999). Four different dialects, namely Tapangu, Tfuya, Lhtu and Iimucu, were spoken by Northern Tsou people. Iimucu is already extinct. The other three dialects are still spoken by Northern Tsou people, though they are all endangered. Most of the Northern Tsou people live in seven adjacent regions namely, Pnguu, Tapangu, Lalauya, Niae'ucna, Saviki, Sinvi and Cayamavana. The Southern Tsou people mainly live in two places San-min and Tao-yuen, which are somewhat to the south of the seven Northern Tsou regions. Two different dialects are spoken by Southern Tsou people, namely Kanakanavu and Sa?alua. These two southern Tsou dialects, according to Li (1999), are almost extinct. The hierarchical relations of the entire Tsou tribe and the languages they speak are illustrated by Figure 5 (Chang 2004).

**Figure 5: Current Tsou Subgroups and Dialects**





### 2.2.2. Southern Min and Hakka

Southern Min makes up about 73.2% of the population, and Hakka about 12% (Chun 2005). Both groups are ethnically Chinese. Most of the current Southern Min and Hakka in Taiwan are the descendents of the settlers who emigrated from the provinces of Fujian (Hokkien) and Guangdong (Canton) (c.f. Figure 1) in Southern China to Taiwan from the seventeenth century until the Japanese colonisation in 1895. Save for some coastal areas, the lands in the two mainland provinces are mountainous. Hence, in order to seek better agricultural lands the people in these two provinces emigrated to Taiwan, Hong Kong and Southeast Asia. The Southern Min and Hakka migrants mainly went to Taiwan in three waves caused by different historical factors. A more detailed description of these waves of migration will be provided in section 2.3.

### 2.2.3. Mainlanders

The mainlanders (also ethnically Chinese) are distinct from the Southern Min & Hakka in terms of the place of origin and the time of their arrival in Taiwan. Chiang Kai-shek headed the KMT (Kuomintang) Party which lost the Chinese Civil War (1945 –1949); he fled to Taiwan with his troops and their families from 1947 - 1949. According to Liao (2000), the great majority of Chiang's troops were from Chinese provinces other than Fujian and Guangdong (though some were from these two provinces), and the languages they spoke varied. This group of immigrants driven by political factors are often called "mainlanders" by the earlier Southern Min and Hakka immigrants. The government-in-exile based in Taiwan led by the KMT Party is referred to as "The mainlanders' regime". According to Chun (2005), the mainlanders make up 13% of the total population of Taiwan today.

## 2.3. Three Waves of Immigration to Taiwan

### 2.3.1. The First Wave of Immigration: The Rule of General Zheng, Cheng-Kong

(1662 – 1683)

Zheng, Cheng-Kong (1624 - 1662) was originally a general of the Min dynasty (1368 - 1644), a regime established by Han Chinese in Mainland China, and was conquered by the Qing dynasty (1644 - 1912) a non-Han Chinese regime. After the collapse of the Min dynasty, General Zheng withdrew his troops (about thirty thousands people, including soldiers and their families) to Taiwan, intending to use it as a military base to fight against the Qing dynasty. According to Kho (2000), Zheng's withdrawal caused the first wave of immigration. Most of these immigrants were Southern Min and Hakka.

### 2.3.2. The Second Wave of Immigration: The Rule of the Qing Dynasty

(1683 – 1894)

The regime established by General Cheng lasted twenty-two years. After his death in 1663, his successors were defeated by the Qing dynasty in 1683. Taiwan then was controlled by the Qing dynasty. Kho (2000) argues that because of the fertile lands and prosperous economy compared to the mountainous and poor lands in the Fujian and Guangdong provinces, many immigrants from these two provinces arrived in Taiwan in search of a better life. During the Qing Dynasty's some two hundred years rule of Taiwan, the population increased rapidly and reached twenty-five million (mostly Southern Min and Hakka) before the Japanese colonisation starting in 1895 (Kho 2000).

### 2.3.3. The Third Wave of Immigration: The Defeat of the KMT (Kuomintang) Party in the Chinese Civil War (1945 –1949)

The third wave of immigration to Taiwan was caused by the Chinese Civil War and the defeat by the Communist Party of the KMT Party led by General Chiang Kai-shek , and his flight to Taiwan with his troops. It was estimated that about one million people arrived in Taiwan from 1947 to 1949 (Kho 2000). Since this war was nationwide, Chiang's soldiers were from twenty-three different provinces of China. The languages, culture, and custom of each province were distinctive. On the one hand, their arrival increased the linguistic and cultural diversity; on the other hand, many conflicts occurred between them and the older immigrants, namely the Southern Min and Hakka people (c.f. section 2.4.). Even today, these conflicts still heavily influence politics in Taiwan.

## 2.4. The Colonisation of Taiwan by Japan (1895 –1945)

### 2.4.1. Education and Language Policies --- Assimilation and Japanisation

Japan's colonisation of Taiwan started in 1894 after the Chinese Qing dynasty's failure in the Sino-Japanese War and ended after Japan lost WWII in 1945. The fifty years of Japanese rule had significant effects upon Taiwan in a number of ways. For instance, much of the infrastructure, including a railway system, ports, and airports, was built by the Japanese government and greatly improved Taiwan's economy. However, as Huang (2000, p.141) suggests, after the Japanese government took control of Taiwan, it "broke off ties between Taiwan and China and cut off Chinese immigration to Taiwan...". More importantly, strict education and language policies were implemented to cut Taiwanese people's cultural links to Mainland China.

In fact, the main purpose of the implementation of education and language policies by the Japanese colonial government was to make Taiwan as Japanese as possible. This point is shown in the content of a public speech delivered by Kodama Gentaroo, the

fourth Japanese Governor-General of Taiwan, in November 10<sup>th</sup>, 1901 (cited by Yide 1937, p.321-322):

“...The ultimate goal of education in Taiwan is to ensure that all Taiwanese people are able to speak Japanese... The policy that Japanese is the only medium of education at all levels of educational institutions will not change... It is crucial that all Taiwanese people are able to speak Japanese because:

- (1) Different varieties of Chinese are spoken in Taiwan, and the language of each Austronesian aboriginal tribe varies. For the convenience of government and communication, the use of a more advanced and superior language, namely Japanese, is therefore necessary.
- (2) Those backward languages, including all Chinese and Austronesian aboriginal languages, should be abandoned...Being able to speak Japanese is a means of assimilating Taiwanese people into Japanese culture and social customs....”

The description of Chinese and aboriginal languages as ‘backward’ and Japanese as ‘advanced and superior’ reveals Governor Kodama’s attitude toward Chinese and Austronesian languages and his intentions to eliminate local cultures in Taiwan. To achieve this goal, educational institutions at various levels, such as elementary schools, junior high schools, high schools and universities, were established in all the major cities in order to educate Taiwanese students about Japanese language, history, religion, and culture. Another government-operated institutions, namely the *Kokugo Denshujō* ‘national language (Japanese) schools’, were also established to teach the general public Japanese (Suemitsu 2004).

In 1940, the Japanese government implemented *Kominka* ‘the Japanisation Movement’. Suemitsu (2004, p.313) points out that the purpose of *Kominka* was to “transform all Taiwanese people to real Japanese through the appreciation of the grace of the emperor, the experience of Japanese social customs, and most importantly the use of Japanese language in daily life”. Some regulations issued by Taipei City Council during the implementation of *Kominka* are shown below:

“ To achieve of the goal of complete Japanisation and assimilation, more *Kokugo Denshujō* ‘national language (Japanese) schools’ in each administrative district will be established, and all citizens in Taiwan:

- a) shall use the national language (Japanese) in public places
- b) shall use the national language in all kinds of social settings
- c) are encouraged to use the national language at home”

(Regulations for Japanese Citizens; cited from Hsu 2004, p.1437)

Suemitsu (2004) also argues that in addition to the implementation of strict language policies, the Japanese government intended to transform Taiwanese people into so-called real Japanese people by other means. For example: Taiwanese people were forced to practice the Japanese Shinto religion and abandon their own religions (e.g., Taoism). No reports or articles written in Chinese were allowed in newspapers. Traditional Taiwanese customs were to be replaced with Japanese ones. Taiwanese people were recommended to use a Japanese name and abandon their original Chinese name. The Japanese government did face some resistance from Taiwanese people when implementing those policies. However, by the use of their superior military powers, the policy of assimilation and Japanisation in Taiwan achieved a certain

degree of success. According to the report of the Education Bureau of the Office of Japanese Governor-General of Taiwan, more than sixty percent of the Taiwanese population were fluent Japanese speakers by the end of 1943, two years before the end of Japanese colonisation (Suemitsu 2004, p.441).

#### 2.4.2. The Overall Impacts of Colonisation by Japan on the Taiwanese People Today and the Current Linguistic Situation

Because of fifty years of language contact and the implementation of strict education and language policies, many Japanese words have been borrowed by Taiwanese languages. Even today, it is not uncommon to find Japanese loan words in the utterances produced by current Southern Min, Hakka or Tsou speakers. However, I argue that the greatest impact of Japan's colonisation on the Taiwanese people was not the introduction of loan words into the Taiwanese languages, but the gradual loss of the people's identification with their original Chinese or Austronesian cultural values. In his diary, Wu Xin-rong, a famous Taiwanese writer from the Japanese colonial period, wrote:

“Everyday we wear Japanese clothes, and eat Japanese food. We feel proud that we have a Japanese style house. We speak Japanese and write in Japanese and even think like Japanese people. It is necessary and convenient for us to do so because we live in a Japanese dominant society. ‘Convenience’ and ‘necessity’ become two major factors, which lead us to the road of assimilation and Japanisation. We, Taiwanese, are a group of people who are forced to assimilate into Japanese culture... Anyone, even the Japanese people themselves, would think we are Japanese.”

(Wu 1938, pp.62-63)

Wu's description reveals the fact that after half a century of colonisation by Japan, many Taiwanese people were confused about their own identity. On the one hand, their original cultural links to Mainland China were cut off; on the other hand, they resisted recognising themselves as Japanese, though their resistance became weaker and weaker as time went by. This situation gradually changed, however, after the end of Japanese colonisation in 1945 and the various conflicts with the later KMT regime led by the mainlanders. Tse (2000) argues that the experience of Japanese colonisation enabled Taiwanese people to form a new kind of identity which allowed them to distinguish themselves from the mainlanders who came to Taiwan after WWII and the Chinese people of communist China.

Although Tse's argument may no longer apply to all Taiwanese people, at least it is still true in relation to a considerable amount of people, in particular the members and supporters of the political party DPP (Democratic Progressive Party). The DPP is a party established by the victims and their relatives of the KMT government's massacres and oppressions (c.f. section 2.5.1) in 1986. Its major political aim is to oppose the mainlanders' KMT government and to claim Taiwan's political independence from the Communist People's Republic of China (PRC). Local Taiwanese languages, especially Southern Min, are used as symbols to represent the newly formed local Taiwanese identity and to distinguish themselves from mainlanders and the Chinese people in the PRC. Hence, I argue that language use by current Taiwanese people is still strongly influenced by the newly formed Taiwanese identity which arose as a result of Japanese colonisation.

## 2.5. The Rule of the KMT Party (1945 –2000)

### 2.5.1. A Brief Review

Taiwan returned to Chinese rule under the KMT Party after Japan's surrender in the Second World War in 1945. At the same time, the Chinese Civil War (1945 – 1949) broke out on the mainland. As stated above, the KMT Party led by Chiang Kai-shek lost all its territories on the mainland and fled to Taiwan. This government-in-exile based in Taiwan imposed Martial Law in 1949, and treated Taiwan as a temporary military base to fight against the Communist Party on the mainland. However, as Huang (2000, p.141) argues, “the KMT government treated Taiwan as a colony, plundering its relatively prosperous economy to pay for a civil war against Chinese Communists on the mainland... Under Chiang, relentless repression and a reign of terror, vocational and residential segregation of Mainlanders and natives, and social, economic and language discrimination alienated the natives and effectively turned Taiwan into a dual society.” Hence, although Taiwan had been returned to Chinese rule, the Austronesian aborigines and the earlier immigrants i.e., Southern Min and Hakka still suffered from the severe exploitation of the KMT government established by the mainlanders. The use of Japanese was entirely forbidden, and the use of Chinese dialects other than Mandarin (e.g., Southern Min and Hakka) in public places was prohibited. As for the Austronesian aborigines, the KMT government forced them to use Chinese names instead of their original Austronesian names; this was part of the plan to eliminate the use of their mother tongue (Liao 2000).

The initial stage of the Chinese KMT government's rule was rather chaotic. Various conflicts between the mainlanders and local Taiwanese people, Southern Min, Hakka and Austronesian aborigines, arose from differences in their culture and social customs. The KMT government's discriminative policies made the situation worse. As



Chen (2006, p.486) argues, “Japanese colonialism was replaced by Chinese nationalism. Either the local Taiwanese or the Austronesian languages and cultures were treated as backward and inferior by the government officials.” To protest against the KMT government’s discriminatory policies, a revolt led by members of the Taiwanese elites who had received Japanese higher education occurred on February 28<sup>th</sup>, 1947. The KMT government used military force to terminate the protest. According to Kho (2000), as a result of the protest over 20,000 members of the Taiwanese elites, mainly doctors, lawyers and similarly educated people, were arrested and massacred by the KMT government. Because of this massacre and the KMT government’s discriminatory policies, the conflicts between the mainlanders and the Southern Min and Hakka peoples, became more intense. Not only did the massacre and conflicts worsen the relationship between the KMT government and local Taiwanese people, but also caused the KMT government to continue its rigid language and social policies.

On Chiang Kai-shek’s death in 1964, his successor Chiang Ching-Kuo took control of the government. Chiang Ching-Kuo’s initiated a series of economic projects to build a comprehensive infrastructure for the development of industry, transportation, and technology, which resulted in two decades of economic boom. Economic prosperity led to the demand from the general public for more socio-political equality. As the demand for a more democratic society intensified, Chiang Chin-Kuo abolished Martial Law in 1987. The government’s policies also became more moderate and egalitarian.

### 2.5.2. Language Policy

During its fifty-five-year rule, the aims of the KMT government’s language policy

changed in response to international and domestic factors. Three distinct approaches to language policy can be identified in three different periods (Chen 1998, pp. 59-78):

- (1) The eradication of Japanisation and the recovery of the Chinese connection  
(1945 –1969)
- (2) The National Language Movement and the consolidation of national identity  
(1970 –1986)
- (3) Multi-cultural and multilingual Policies (1987 – 2000)

#### 2.5.2.1. The Eradication of Japanisation and the Recovery of Chinese Connection (1945 –1969)

At the start of the KMT party's rule, Japanese was very frequently used by Taiwanese people. The main goals of the government's policies in this period were the eliminating of any effects caused by Japanese colonisation, "the promotion of the National Language (Mandarin) Movement<sup>3</sup> and the recovery of Taiwanese people's connection toward Chinese culture and customs" (Chen 1998, p. 60). The publication of Japanese language materials was forbidden. Punishment and fines were imposed on the use of Japanese and languages other than Mandarin in public places. Some regulations related to the promotion of the use of Mandarin issued by the government of Taiwan Province, R.O.C during 1945 to 1969 are shown below:

"1. Any form of speech or announcement at all levels of educational institutions should be delivered in the national language (Mandarin).

---

<sup>3</sup> According to Tse (1986, p.25), "The selection of Mandarin Chinese as the national language and the subsequent standardization and propagation process which began in China in the early 1910s have officially been called the National Language Movement (NLM). The NLM in Taiwan started in 1946".

2. Teachers and students of all levels of educational institutions should only use the national language for communication. Any violation of this regulation shall be punished according to relevant regulations.
3. Any materials written in Japanese, including comic books, books, documents etc, should be destroyed.
4. The promotion of the use of the national language in all regions where the Austronesian aborigines live should be stricter.
5. The national language should be the only medium of education in the primary schools. The use of Japanese is forbidden. Austronesian languages may be used as supplements to the national language for the purpose of educating students with lower levels in the elementary schools.
6. The national language should be the only language spoken in all government/private institutions, schools and public places.
7. The broadcasting of films in Japanese or languages other than the national language is strictly forbidden.
8. The translation of foreign films should only be into the national language, but not into any other varieties of Chinese.”

(A summary of Taiwan Provincial Government Release, 1945 – 1969)

The regulations shown above suggest that the major goal of the KMT government's policy was to eliminate the influences of the Japanese colonisation of Taiwan. Although the enforcement of these strict and discriminatory regulations led the Taiwanese people to rise up against the KMT government, they were nevertheless successful in prohibiting the use of Japanese and promoting the use of Mandarin.

### 2.5.2.2. The Continuation of the National Language Movement and the Promotion of Nationalism (1970 –1986)

In the 1970s, the KMT government experienced a serious diplomatic crisis in international politics. Instead of recognising the Republic of China (R.O.C) established by the KMT government, the United Nations (UN) recognised the People's Republic of China (PRC) established by the communist party as one of its permanent members. Most of the countries in the world broke off formal diplomatic relations with the KMT led government of R.O.C. In addition to this diplomatic crisis, the government also faced a series of domestic problems. Many uprisings that occurred from 1945 to 1969 were due to the Taiwanese people's resistance of the government's prohibition on the use of Southern Min, Hakka or Austronesian languages. Hence, as Tse (2000) argues, the promotion of the use of the languages other than Mandarin was viewed as a form of treason. To solve their internal and external problems, the KMT government decided to reinforce the promotion of the National Language Movement in order to consolidate Taiwanese people's national identity. This is illustrated by some of the regulations issued by the government between 1970 and 1986 shown below:

- “1. All levels of educational institutions should open more courses to teach the general public the national language in order to strengthen people's national identity toward our country.
2. The use the national language is a symbol of loyalty toward our country. The staff of all educational institutions should speak the national language in order to be good models for the students.
3. All educational institutions should increase the number of national language,

national history and geography lessons in order to strengthen students' identity toward our country.”

(A summary of the Taiwan Provincial Government Release, 1974 – 1986)

Because of the KMT government's non-egalitarian language policy, Mandarin became politically and socio-economically a more prestigious language. According to Chen (1998), a high level of Mandarin proficiency was a basic requirement for people who wanted to work in the government sectors and educational institutions, or to acquire higher socio-economic status. Furthermore, as Huang (2000) argues, the discriminatory language policy led to a language shift by younger generations of Southern Min and Hakka speakers from their mother tongues towards Mandarin and the endangerment of Austronesian languages such as Tsou.

#### 2.5.2.3. Multicultural and Multilingual Policy (1987 – 2000)

The abolition of Martial Law in 1987 led to more moderate language and socio-economic policies. Liao (2000, p.169) points out that “since the lifting of Martial Law, the government's attitude has changed from that of eliminating local languages to that of promoting them.” Native language teaching programmes, including Southern Min, Hakka, and Austronesian languages, have become compulsory in elementary and junior high schools. The use of Southern Min, Hakka, and the Austronesian languages has also been encouraged in the government sectors, other institutions, and in private enterprises.

Nevertheless, the liberal nature of the language policy also led to greater conflicts among different groups in Taiwan. According to Liao (2000), after the abolition of Martial Law, some political parties seemed to manipulate the issue of language

equality between all languages in order to obtain political support from the public. For instance, in the presidential election in 1996, the candidates of the DPP, whose supporters were mainly Southern Min people, used Southern Min as the main language in a television campaign in order to capture the attention of the Taiwanese people. Thus, Tse (2000, p.161) proposes that the use of the local mother tongue became “an expression of democratisation, a sign of localism and an assertion of ethnolinguistic identity...language seemed to be a dividing force rather than a unifying one, increasing the social and psychological distance among the major ethnic groups”. Although the DPP did not win the 1996 election, it won the second presidential election held in 2000, and ended the KMT party’s fifty-five years of rule.

#### 2.6. The Rule of the DPP Party and Current Language Policy (2000 – 2007)

As mentioned earlier, the DPP party is a party which advocates the independence of Taiwan. Mainland China, the PRC, however treats Taiwan as one of its provinces, and strongly opposes Taiwan’s aim of political independence. Hence, across-strait political conflicts have been much greater since the DPP party came into power in 2000. By advocating anti-China policies and allowing the public to vote on constitutional changes, the DPP leader Chen Shui-bien again defeated the KMT party and its allies, those favouring close ties with China, and won the third presidential election held in 2004, but by a margin of only 0.2% (BBC News January 19, 2006). The result of this election to certain extent reflected the conflicting senses of identity of Taiwanese people. About half of the population see themselves as Taiwanese rather than Chinese, while the other half want to keep close ties with Mainland China. Such a phenomenon is also reflected in their language use. Since the supporters of the KMT party are more dominant in the northern party of Taiwan, Mandarin is more frequently used there. Southern Min is used as the major medium of communication in the south, where

most of the supporters of the DPP party reside.

The current language policy of Taiwan remains the same as that implemented from 1987 to 2000 and aims to create a multi-cultural and multi-lingual Taiwanese society. Many government funded projects have been launched to save the endangered Austronesian languages and to revitalise Hakka. Nevertheless, as Liao (2000) argues, all languages in Taiwan basically have equal status, but Southern Min seems to be gradually acquiring higher status than other languages. This may be because of the ethnic background of the members of the DPP party, mostly Southern Min, and the government's anti-China policies. As Mandarin is also the official language of the People's Republic of China, many DPP members of Parliament have decided to speak Southern Min in the parliament rather than Mandarin, the official language. The Ministry of Education has gradually decreased the amount of classical Chinese, Chinese history and Chinese geography lessons in educational institutions.

Although the status of Mandarin is often downplayed by the DPP government, almost all Taiwanese people are fluent speakers of it due to the KMT governments' language policy. Furthermore, despite the political tension, Taiwan and Mainland China have very close trading relations. The fast-growing economy of Mainland China has attracted many Taiwanese people, regardless of their political positions, to set up businesses in this huge market. Hence, Mandarin remains the most important medium of communication in Taiwan.

## 2.7. Conclusion

Because of the discriminatory language policies implemented by the KMT government in Taiwan in the 1970s and 1980s, Southern Min, Hakka, and the

Austronesian languages were severely suppressed. This caused the loss of younger speakers of Southern Min and Hakka, and the endangerment of the Austronesian languages. Under the influence of this monolingual language policy, comparatively little research in the past focused on the investigation of Southern Min, Hakka, and the Austronesian languages. The current multicultural and multilingual policy provides a good climate to conduct a more thorough investigation of these languages. It is hoped that, in respect to the study of the code-switched utterances in the two language pairs, Mandarin/Southern Min and Mandarin/Tsou, this thesis could show how Mandarin, Southern Min and Tsou are used in these two language communities, and what the future of an endangered language such as Tsou may be.



## Chapter 3: Review of the Literature

### 3.1. Introduction

It is often the case that languages come into contact through bilingualism or multilingualism. People, who are brought up in an environment of two or more languages, or language varieties, may be bilingual or multilingual. These bilingual/multilingual speakers may switch from one language/language variety to another in their utterances. The phenomenon is known as codeswitching<sup>4</sup> (CS will be used in the following discussion). Different linguists propose their own interpretations of this term. This thesis will adopt Myers-Scotton's (2002a, p.3) definition: "CS is defined as the use of two languages in the same clause". I am adopting Myers-Scotton's definition for two reasons. First, one of the major aims of this thesis is to test the MLF model she proposes. Second, although Muysken's (2000) typological approach will also be tested, an important theoretical problem of his model is found, namely there is a lack of basic unit of analysis (c.f. section 6.3.1.). The solution to this problem, which is proposed by Deuchar, Muysken and Wang (paper in preparation), is to use a clause as a basic unit of analysis. Hence, Myers-Scotton's (2002a) definition of CS will be adopted in this thesis.

In Chapter 1, we briefly described the three major perspectives on CS in the current literature, namely sociolinguistic, CA (Conversational Analysis), and structural. This chapter will include a more detailed discussion of various perspectives on CS in current literature. The organisation of this chapter is as follows: First, various perspectives on CS will be reviewed. Second, we shall provide a comprehensive introduction to the two structural models this thesis intends to test, namely

---

<sup>4</sup> The term "code-mixing" is also used by some scholars (e.g., Muysken 2000). We shall use CS (codeswitching) in this thesis.

Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach. The discussion of these two models will include the main arguments of these two models, and an augmented MLF model, the 4-M model. Then, we shall compare these two models, and evaluate them with reference to studies published by other linguists. Finally, we shall review previous CS studies related to Chinese.

### 3.2. Various Approaches to CS

#### 3.2.1. The Sociolinguistic Approach

Various research studies have been published to explain why people switch codes. Sociolinguists try to answer this question with reference to various extra-linguistic factors. For instance, Fishman (1965) proposes several factors including situation and topic, which may influence people's choice of language. Taking the factor "situation" as an example, he argues that one may switch between different languages depending on the situation, for certain languages or a certain style of a language is "more likely to be reserved for certain situations than the other" (p.91). For instance, in a formal conference the use of formal language is usually preferred. If the conference is an international one, English may be preferred because it is a global language. However, during breaks conference participants may use languages other than English, such as French, or German, to talk to other people who share a similar linguistic background. Hence, people switch from one language to another according to the different situations they may be in.

In addition to discussing the role of "situation", Fishman (1965, p.92) also argues that when discussing certain topics, one may switch between different languages because "certain topics are somehow handled better in one language than in another". To

illustrate this point, consider Nartey's (1982) research. In her study of English/Dāŋme CS utterances produced by younger educated Ghanaians, Nartey (1982) argues that extra-linguistic factors play important roles in influencing young Ghanaians' language choice. For instance, she states that much of the western knowledge on subjects like economics, education, politics, and industrial civilization was introduced into Ghana through English during the British colonial period. She found that when those topics were discussed, Ghanaians would switch from the Ghanaian language to English. Moreover, Nartey found that age was an important variable when analyzing English/Dāŋme CS utterances. The results of her study showed that the majority of CS was done by members of the younger generation (approximately the age group of 15-45), whereas the older generations (over 50 years old) kept the two languages apart.

People's CS behaviour may also be triggered by the entire political and socio-economic environment of the language community in question. For example, in her studies of the use of French and English in Ontario and Quebec (Canada), Heller (1992) claims that in the context of the unequal distribution between the political and socio-economic power of Anglophones and Francophones (i.e., Anglophone-dominant), the monolingual use of either English or French reflects an individual's identity as belonging to a particular language group. Hence, she argues that for those who want to access to both the English and French worlds to maintain good social relationships or achieve certain socioeconomic goals, CS is used strategically and symbolically to neutralize the tension between these two groups.

### 3.2.1.1. Evaluation

From previous sociolinguistic research on CS, we know that people's choice of language (or codes) is influenced by a variety of extra-linguistic variables. However, is one's CS behaviour always related to those extra-linguistic variables? Auer (1998, p.3) argues that "macro-sociolinguistic aspects of the speech situation never determine completely language choice, including CS and the absence of it." He goes on to argue that the organization of a conversation and the participants' interaction, which are often ignored by the practitioners of a sociolinguistic approach to CS, are also important factors that may influence people's language behaviour. Furthermore, Myers-Scotton (2001) argues that even if the statistical analysis of sociolinguistic studies show that a great majority of the members of a given community uses a particular language in specific social settings or situations, it is very unlikely that every single member would follow this community norm. Hence, a purely sociolinguistic approach to CS is evidently inadequate to explain why people alternate between languages. Although the sociolinguistic approach to CS will not be adopted to test the Mandarin/Tsou and Mandarin/Southern Min CS data in the present study, the sociolinguistic variables, namely age and educational background, will still be used in the analysis of the data in Chapter 7.

### 3.2.2. Conversational CS

In contrast to a sociolinguistic approach, which examines the relationship between codeswitched utterances and various extra-linguistic factors, linguists such as, Gumperz (1982); Auer (1998; 2005); Li Wei (2002; 2005), look at the relationship between CS and conversational structure, and seek to investigate the communicative functions behind codeswitched utterances in order to have a better understanding of the conversational interaction.

Gumperz (1982) argues that codeswitched utterances signal socio-pragmatic information such as, shared background knowledge of the participants of the conversation, and the use of CS can achieve different communicative functions (e.g., quotations, addressee specification, reiteration, and personalization vs. objectivization). Now, consider Gumperz's examples in (1) and (2).

(1) Hindi-English. Father calling his small son, who was learning to swim in a swimming pool:

*Baju-me jao beta, andar mat* (go to the side son, not inside). Keep to the side.

(Gumperz 1982, p.78)

(2) Hindi-English. College girls talking about what a male friend had said:

a. Speaker A: *Tera nam liya, lipa ka nam liya* (he mentioned you, he mentioned Lipa).

b. Speaker B: *əha kya kəkne* (ah what should I say) *she'll be flattered*. *Ajmāī leke a rəhi thi na* (today I was going to bring her see).

(Gumperz 1982, p.81)

In (1), the speaker reiterates the same message by the use of a CS strategy to remind his son that it is dangerous to leave the side of the pool. Gumperz (1982, p.81) argues that in (2) speaker B switches from Hindi to English in order to suggest that “the statement is a casual one, not implying personal involvement”, and switches back to Hindi for stating what she personally plans to do. He goes on to argue that in example (2), speaker B presupposes that speaker A recognizes the fact that the use of English

represents the speaker's neutral attitude, for both speaker A and B belong to the same social group and share the background knowledge of that group. Hence, those who do not share the same background knowledge cannot understand all of the information carried by codeswitched utterances.

Auer (1984; 1998; 2000) proposes a CA model of CS, which makes a distinction between discourse-related CS and participant-related CS. Discourse-related CS refers to "the use of code-switching to organise the conversation by contributing to the interactional meaning of a particular utterance" (Auer 1998, p.4). In other words, he argues that CS may be triggered by the organisation of the ongoing conversation (e.g., turn-taking, repairing, and topic-shifting). For participant-related CS, the participants in a given conversation "search for an account within the individual who performs this switching or his or her co-participants" (p.8). This means that CS may be triggered because of the participants' or co-participants' "preference or competence" in a certain language (Auer 2000, p.176). Li Wei (2002, p.168) uses his Cantonese/English CS example, shown in (3), to illustrate the notion of discourse-related and participant-related CS.

(3) Speaker A is a man in his late twenties. Speaker B is a 40-year-old woman.  
Speaker C is Speaker B's teenage daughter.

a. Speaker A: *Sik gai a*  
eat chicken part.  
'Have some chicken.'

b. Speaker B: mm.  
(5.0)

c. Speaker A: Haven't seen XXX (name, three syllables) for a long time.

(2.0)

- d. Speaker A: Have you seen him recently?
- e. Speaker B: No.
- f. Speaker A: Have you seen XX (name, two syllables)?
- g. Speaker B: (2.0) (To Speaker C) *Ning ngaw doi haai lai.*  
Bring my those shoe ASP.  
'Bring my shoes here.'
- h. Speaker B: (To Speaker A) *Koei hoei bindon a?*  
he go where PA  
'Where was she going?'

The conversation in (3) is started by Speaker A in Cantonese. The five-second silence in (3b) then indicates the end of the interaction. Speaker A then switches to English and attempts to introduce a new topic. However, Speaker B gives no response, which is signalled by the two-second silence in (3c). A continues the same topic and turns his utterance into a question. Then, B answers A's question briefly in English. A then continues to use English to ask about a different person. B does not respond to A's question, but speaks to C in Cantonese. Finally, B turns back to A and speaks in Cantonese.

Li Wei argues that the conversation in (3) indicates that English is not preferred by Speaker B. For instance, after a two-second silence B does not respond to A's utterance in (3c). Even if B uses English to answer A's question in English, as shown in (3e), her answer is negative and minimal. Hence, Li argues that (3g) could be an example of participant-related CS, for B's language alternation is a result of her preference for Cantonese.

An example of discourse-related CS would be A's switch to English in (3c). The entire conversation starts with Cantonese used by A. After a five-second silence in (3b), A decides to introduce a new topic, so he switches to English. (3g) is also an example of discourse-related CS. A's question in English in (3f) and the two-second pause in (3g), together, form an adjacency pair. Li Wei argues that B uses the two-second silence to soften the impact of the change of addressee (request to C) and language (to Cantonese). When she returns to A's question in the next turn in (3h), "Cantonese appeared to be more legitimate because it followed from the immediately preceding turn" (p.169). Li Wei argues that B uses the nature of conversational organization (i.e., turn-taking) as a strategy to switch to the language she prefers.

#### 3.2.2.1. Evaluation

The turn-by-turn and sequential analysis of the CA approach to CS certainly provides very interesting insights into people's language choice. However, a CA approach to CS is sometimes criticised for paying too much attention to the organisation of a conversation. For instance, Myers-Scotton (2001) argues that CA practitioners often ignore the wider socio-political context. As discussed earlier, sociolinguists have established a relationship between people's code choice and a variety of extra-linguistic variables (e.g., social settings, an individual's social network, or the socio-economic environment of the community). Some CA practitioners, such as Auer (1998), claim that a CA approach does link codeswitched utterances to some extra-conversational factors. For instance, he argues that preference-related CS often signals a preference for a certain language which may have been formed by the socio-political environment of the speech community in question. However, members of the same community may acquire a given language or produce CS utterances for a



variety of reasons. Hence, it is questionable whether any one analyst's interpretation of CS utterances will be accepted consistently by other analysts or other members of the same community.

The other problem the CA approach has, is that the distinction between discourse-related and participant preference-related CS is not clear. For instance, I argue that speaker B's use of Cantonese in (3g) can be interpreted not only as a preference-related CS, but also as a discourse-related CS (i.e., topic-shifting). One of the earliest practitioners of the CA approach to CS, Auer (1998) also agrees that there is no clear distinction between these two notions. Hence, if the basic distinction between discourse-related and participant-related CS is not clear, then the interpretations of CS utterances with reference to these two notions are likely to be confusing. Since the second aim of the present study is to predict the future of an endangered language, Tsou, the information that the CA approach to CS could provide is more closely related to the speakers (e.g., switching between codes to achieve the speaker's purposes in communication) rather than the language itself. For this reason, the CA approach will not be adopted to analyse the two language pairs (i.e., Mandarin/Tsou and Mandarin/Southern Min) in this thesis.

### 3.2.3. A Structural Approach to CS

It is clear that the sociolinguistic and CA approaches consider interesting aspects of CS phenomena. Both approaches provide possible answers to the question: why do people switch codes? The sociolinguistic approach attempts to explain to what extent extra-linguistic variables may influence people's switching between codes. With the adoption of a sequential analysis, CA practitioners attempt to investigate why CS occurs at a particular point in a given conversation and what communicative functions

or social meanings are carried by CS.

In contrast with the sociolinguistic and CA approaches to CS, linguists who adopt a structural approach aim to investigate how people switch codes. In the history of CS research, various grammatical constraints have been proposed to show that CS utterances are rule-governed and do not occur randomly. The discussion below will provide an overview of some of the grammatical constraints proposed in the literature.

### 3.2.3.1. Poplack's Free Morpheme and Equivalence Constraints

Among various grammatical constraints that have been proposed, Poplack's (1980) **free morpheme constraint** and **equivalence constraint** have been among the most influential. These were based on her research in the bilingual (English/Spanish) Puerto Rican community.

#### *The Free Morpheme Constraint*

Poplack's (1980, p.595) definition of the free morpheme constraint is provided below:

#### *"The Free Morpheme Constraint:*

*Codes may be switched after any constituent in discourse provided that the constituent is not a bound morpheme...This constraint holds true for all linguistic levels but the phonological ..."*

To illustrate this constraint, consider her English/Spanish examples in (4) and (5). The English word "excuse" in (4) is a free morpheme, and therefore it can occur as a switched element.

(4) una buena *excuse*

‘a good excuse’

(Poplack 1980, p.586)

In (5), “eat” is an English free morpheme, “-iendo” in Spanish is a bound morpheme.

According to the free morpheme constraint, such a switch is unlikely to occur.

(5)\* eat + iendo

‘eating’

(Poplack 1980, p.586)

### *The Equivalence Constraint*

Poplack’s (1980, p.585-6) definition of the equivalence constraint is provided below:

#### “The Equivalence Constraint:

*Code-switches will tend to occur at points in discourse where juxtaposition of L1 and L2 elements does not violate a syntactic rule of either language, i.e. at points around which the surface structures of the two languages map onto each other. Thus, a switch is inhibited from occurring within a constituent generated by a rule from one language which is not shared by the other.”*

Consider her example in (6):

(6) a. English:	I	told him	that	so	that	he	would bring it	fast
	↑	↙ ↘	↑	↑	↑		↙ ↘ ↗	↑
b. Spanish:	(Yo)	le dije	eso	pa’	que	(él)	la trajera	ligero
c. CS	I	told him	that	[pa’que			la trjera	ligero].

(Spanish/English; Poplack 1980, p.586)

She argues that the vertical lines in (6) indicate the permissible switch points and the arrows show how the surface structures of English and Spanish map onto each other. The invented Mandarin/Tsou utterance shown in (7c) would be inhibited according to the equivalence constraint. As shown in (7a), the demonstrative *na* ‘that’ occurs before the noun *xiao-hai* ‘child’. However, the Tsou demonstrative *tonoi* ‘that’ appears after the noun *oko* ‘child’. The surface structure of the DP ‘that child’ in Mandarin and Tsou is not equivalent. Therefore, a Mandarin/Tsou CS utterance like (7c) is unlikely to occur.

(7) a. *na xiao-hai* (Mandarin)

that child

b. *oko tonoi* (Tsou)

child that

c.\* *na oko* (invented Mandarin/Tsou CS example)

that child

CS data in different language pairs have been reported as counter-examples of Poplack’s two constraints by other linguists. For instance, in her English/Adāŋme CS data, Nartey (1982) found numerous counter-examples to Poplack’s two constraints. Consider her example in (8).

(8) a    ηε    mī    *helpe*

they copula me help (present progressive)



are isolating languages, they only have a rather limited number of free morphemes, such as the aspect markers. Hence, the free morpheme constraint almost does not apply to the Mandarin/Southern Min CS data.

However, many violations were found when applying Poplack's model to the Mandarin/Tsou CS data. Since the basic word order of is SVO, while that of Tsou is VOS, the differences of their surface word order yielded many counter-examples to the equivalence constraint. Consider my bilingual Mandarin/Tsou example in (10a) and its monolingual version in Mandarin and Tsou in (10b) and (10c).

(10) a. iu?fafoinana *de* *xiang-fa* <sub>subject</sub> *bu* *yi-yiang*.

Young people poss. thought neg. same

'Young people's thought was different.'

(normal print = Tsou; Italic =Mandarin)

b. mo a?hto mɛh?mɛskə ?e totohɛgɛ

tense2,Agent neg. resemble Nom1 thought

da iu?fafoinana <sub>subject</sub>(Tsou)

Obl1 young people

'Young people's thought were different.'

c. *nian-qing-ren* *de* *xiang-fa* <sub>subject</sub> *bu* *yi-yiang*. (Mandarin)

Young people poss. thought neg. same

'Young people's thought were different.'

(10a) can be regarded as a clear counter-example of the equivalence constraint for two reasons. First, the noun *iu?fafoinana* 'young people' in the monolingual Tsou example in (10b) appears in the sentence final position, whereas the noun *nian-qing-ren* 'young

people' in the monolingual Mandarin sentence in (10c) occurs in the sentence initial position. Hence, the occurrence of the switched element *iu?fafoinana* 'young people' in the bilingual Mandarin/Tsou example in (10a) clearly violates the syntactic rule of Tsou, and therefore violates the equivalence constraints.

Second, as shown by the monolingual examples in (10b) and (10c), there is a difference in terms of the surface word order of the linguistic elements in the underlined noun phrases. In Tsou, the noun *iu?fafoinana* 'young people' should occur after the noun *totohəgə* 'thought', whereas in Mandarin the noun *nian-qing-ren* 'young people' should appear before the noun *xiang-fa* 'thought'. According to the equivalence constraint, no switch is allowed in the noun phrase since the syntactic rules of the two languages are different. Hence, the bilingual example in (10a), which is collected from my Mandarin/Tsou corpus, is a clear violation of the equivalence constraint. In fact, because of the typological difference of Mandarin and Tsou, many examples resembling (10a) were found in the Mandarin/Tsou corpus. The examples proposed by many other linguists e.g., Nartey (1982) as well as by many of my Mandarin/Tsou data reject Poplack's model. Furthermore, both Mandarin and Southern Min are isolating languages in which most of the morphemes are free morphemes. The free morpheme constraint was found not to be applicable to almost all of the Mandarin/Southern. For these two reasons, Poplack's model was not adopted by the present study.

### 3.2.3.2. The Government Constraint

DiSciullo, Muysken, and Singh (1986) also criticize Poplack's constraints, and argue that the equivalence constraint would only be applicable if languages A and B have equivalent categories. They go on to argue that if language A has the categories, such

as determiner or conjunctions, then language B must also have these two categories, otherwise, it is unlikely to determine whether a switching point would occur between the determiner and conjunctions. However, many categories are found only in specific languages (e.g., the classifiers in Chinese are not found in English), so a category in language A may not totally correspond to one in language B. Hence, the equivalence constraint would fail to predict some of the switching points in these languages.

Furthermore, they point out that Poplack merely looks at the linear sequence of the codeswitched utterances, but ignores the hierarchical aspect of CS. According to DiSciullo, Muysken, and Singh (1986, p.3), “most grammatical principles are formulated in terms of hierarchical relations rather than of linear order and since CS appears to involve central aspects of grammatical competence... it would be necessary for a grammatical constraint on CS to be structural rather than linear”.

Rather than the constraints proposed in Poplack’s theory, they argue that “the process of CS was in fact constrained by government relations” (p.1). Before outlining their model, it is necessary to describe briefly what government relations mean. The definition of government formulated by Chomsky (1981, p.164) is provided as below:

**“ $\alpha$  governs  $\Upsilon$  in [ $\beta \dots \Upsilon \dots \alpha \dots \Upsilon$ ], where:**

**(i)  $\alpha = X^0$**

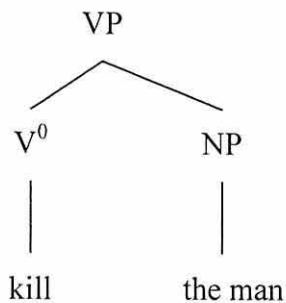
**(ii)  $\alpha$  and  $\Upsilon$  are part of the same maximal projection”**

According to Chomsky’s definition, it is the head ( $X^0$ ), which governs its complement

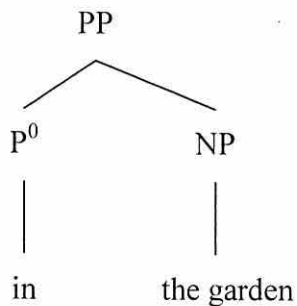


(Y) within the same maximal projection. Consider the examples in (11) and (12). In the verb phrase in (11), the  $V^0$  head “kill” governs the NP “the man”. In the prepositional phrase in (12), the  $P^0$  head “in” governs the NP “the garden”.

(11) I [kill the man]<sub>VP</sub>.



(12) I met him [in the garden]<sub>PP</sub>.



DiSciullo, Muysken, and Singh (1986, p.5) applied government theory to propose a grammatical constraint in CS, and argued that “if X has language index q and if it governs Y, Y must have language index q”. This statement was turned in to the formula shown as (13).

(13) If X governs Y, ... then  $Xq \dots Yq \dots$

(DiSciullo, Muysken, and Singh 1986, p.5)

Also, consider the invented Mandarin/English CS examples in (14) and (15).

(14) I [ kan le ( na ben shu )<sub>NP</sub> ]<sub>VP</sub>

I read asp,perf that Mclass2 book

‘I have read that book.’

(15) \*I had my lunch [ in ( na can-ting )<sub>NP</sub> ]<sub>PP</sub>

I had my lunch in that restaurant].

‘I had my lunch in that restaurant.’

According to DiSciullo, Muysken, and Singh’s model in (13), an English-Mandarin CS utterance like example (14) is permitted. This is because the verb *kan* ‘read’ is the head of the underlined verb phrase, and governs its complement, namely the NP *na shu* ‘that book’. Hence, the verb and its complement NP should be in the same language. The example in (15), however, is not allowed. The preposition “in” is the head of the underlined prepositional phrase (PP), and it governs its complement NP, namely *na can-ting* ‘that restaurant’. Since these two are in different languages, an example like (15) should not occur.

However, many examples have been found which are counter to DiSciullo, Muysken, and Singh’s government model. Consider their example in (16).

(16) ha ricevuto [il *diplôme*]<sub>NP</sub>

‘has received the diploma’.

(Italian/French; DiSciullo, Muysken, and Singh 1986, p.13)

According to the government constraint, the Italian verb “ricevuto” in example (16) governs the NP - [il *diplôme*]. Hence, this governed NP should also be Italian. However, in example (16), the CS occurs in the governed NP which clearly violates the constraint. Sciullo, Muysken, and Singh then introduce a neutralizing element: the **language index carrier**. They argue that ‘the highest lexical element in a maximal projection determines the q index of the projection and such an element is called the **Lq carrier- language index carrier**’ (DiSciullo, Muysken, and Singh 1986, p.6). They go on to claim that “If Lq carrier has language  $q$ , then  $Y^{\max}_q$ ”, which can be interpreted as: “the first element of the constituent will decide the language index of the entire constituent” (p.6). With reference to this claim, the Italian determiner “il” in example (16) will carry the index “Italian” for the whole NP even though the N is in French.

Nevertheless, such a claim itself is incompatible with the definition of government – the noun head should govern the whole NP. Moreover, if the language index carrier decides the language index of the whole constituent, switches of adverbs in a VP should never occur. Nevertheless, many counter-examples, such as (17), are found.

(17) La [lascia *toujours* sulla tavola.]<sub>VP</sub>

‘She leaves it always on the table.’

(Italian/French; DiSciullo, Muysken, and Singh 1986, p.15)

Thus, a more limited definition of government is required. In other words,

government only refers to the relationship between the lexical head and its complement, but modifiers should be excluded.

One may argue that in example (16) it is the determiner “il” which governs the noun “*diplôme*” (therefore a DP), especially when the notion that functional categories, such as determiners, quantifiers, complementizers, and auxiliaries are the heads, becomes more widely accepted. Similarly, an auxiliary or finite tense marking on the main verbs is often seen as governing the subjects and selecting a participle and complement verb (Muysken 2000). However, DiSciullo, Muysken, and Singh’s (1986) Italian/French examples in (18) and (19) are clearly in contradiction to this view.

(18) [*La plupart des canadiens*]<sub>subj</sub> scrivono ‘c’.

‘The majority of Canadians write ‘c’.’

(Italian/French; DiSciullo, Muysken, & Singh (1986, p.15))

(19) No, *parce que* hanno *donné des cours*.

‘No, because they gave lectures.’

(French/Italian/English; DiSciullo, Muysken, & Singh (1986, p.15))

Following the discussion above, a different version of government constraints is proposed and shown in (20).

(20) \* [Xp Yq], where X L-marks<sup>5</sup> Y, and p and q are language indices.

---

<sup>5</sup> The definition of L-marking proposed by Chomsky (1986) is:

A head  $\alpha$  L-marks  $\beta$  if  $\alpha$  is lexical and  $\theta$ -governs  $\beta$ .

$\alpha$  is the lexical head (N,V,A,P) which assigns a  $\theta$ -role to  $\beta$ . In other words,  $\beta$  is the immediate

(Muysken 2000, p. 23)

This revised version of government constraints refers to the relationship between the lexical heads (e.g., N, V, A, P but not functional heads like determiners and auxiliaries) and their immediate complements (Muysken 2000, p. 23). Thus, if X L-marks Y, then these two should have the same language index. However, even with this revised version, the government constraint is still problematic. A number of switches, which occur between the verb and its direct object, have been reported. For instance:

(21) *Žib li-ya een glas water of zo.*      V IO/DO<sup>6</sup>

‘Get for-me a glass of water or so.’

(Muysken 2000, p.24, citing Moroccan Arabic/Dutch; Nortier 1990, p. 131)

(22) *Anaka-ndir intercultureel werk.*      V/DO

‘I I-am-doing intercultural work.’

(Muysken 2000, p.24, citing Moroccan Arabic/Dutch; Nortier 1990, p.131)

In view of the counter-examples discussed in this section, the government constraint was abandoned by Muysken (2000).

### 3.2.3.3. A Minimalist Approach to CS

Macswan (2000; 2005) adopts Chomsky’s (1995) minimalist programme in which parameters are restricted to the lexicon and grammars are lexically-encoded to examine CS phenomena. Since CS often includes lexical items from two different

---

complement of  $\alpha$  but not an adjunct.

<sup>6</sup> “V” refers to verb. “IO” refers to indirect object, and “DO” refers to direct object.

languages, Mascan (2005, p.5) argues that “CS is formally the UNION of two (lexically-encoded) grammars”. According to the minimalist programme, PF (phonetic form), namely the phonological elements that are actually uttered, contains all the rules and constraints. Hence, a PF element of lexicon A (PF<sub>1</sub>) contains certain grammatical rules. A PF element of lexicon B (PF<sub>2</sub>), however, contains different grammatical rules. CS is a union of two lexicons, which contains only parts of the grammars of language A and B, and therefore cannot meet the requirement imposed by PF<sub>1</sub> or PF<sub>2</sub>. Therefore, Macswan (2000, p.45) argues that “CS at the level of PH is not possible”. His argument can be summarised below.

#### “PF Disjunction Theorem

- (i) The PF component consists of rules/constraints which must be (partially) ordered/ranked with respect to each other, and these orders/rankings vary cross-linguistically.
- (ii) Code switching entails the union of at least two (lexically encoded) grammars.
- (iii) Ordering relations are not preserved under union.
- (iv) Therefore, code switching within a PF component is not possible.”

(Macswan 2000, p.45)

Macswan’s (2000; 2005) PF Disjunction Theorem predicts that switching of phonological systems between a stem and an affix, which is formed internal to the lexicon, is prohibited. A phrasal affix, such as the possessive “’s” in English which are not formed internal to the lexicon, is therefore not affected. To illustrate these points, consider his examples in (23).

(23) \* a. Juan está eat-iendo.

Juan be/3Ss eat-DUR.

‘Juan is eating.’

(Spanish/English; Macswan 2005, p.6; citing Macswan 1999, p.222)

b. Juan está parqueó su coche.

Juan be/3Ss park-DUR his car

‘Juan is parking his car.’

(Macswan 2005, p.7)

(3Ss = third person singular subject agreement; DUR =durative aspect)

c. Su novia ’s coche está nuevo.

His girlfriend-GEN car is new

‘His girlfriend’s car is new.’

(Spanish/English; Macswan 2005, p.14; citing Hlavac 2003, p.165)

According to the PF Disjunction Theorem, the example in (23a) should not occur. This is because the underlined element “*eat-iendo*” has an English stem “eat” which still preserves English pronunciation, and has a Spanish durative aspect suffix “-iendo” which has Spanish pronunciation. In other words, the underlined element “*eat-iendo*” contains two different phonological systems. According to the PF Disjunction Theorem, (23a) is unlikely to occur. Macswan argues that the occurrence of (23b) is possible. This is because the underlined element in (23b) contains only one phonological system (i.e., Spanish), for the English stem “park” is borrowed and is phonologically integrated into Spanish. Moreover, a codeswitched utterance like (23c) is possible. The English possessive marker “’s” is a phrasal affix and therefore is not prohibited by the PF Disjunction Theorem.

Although Macswan's PF Disjunction Theorem is formulated based upon Chomsky's (1995) minimalist programme, it overlaps Poplack's (1980) free morpheme constraints in terms of their empirical prediction of the occurrence of CS. Hence, the counter-example to the free morpheme constraint reported by Nartey (1982) can also be treated as a counter-example to the PF Disjunction Theorem. Consider Nartey's (1982) can example in (8), repeated as (24) below.

- (24) a    ηε       mī    helpe  
          they copula    me    help (present progressive)  
          ‘They are helping me.’  
  
          (English/Adāηme; Nartey 1982, p.185)

In example (24), the underlined suffix is an Adāηme present progressive marker “e”, while Nartey reported that the stem “help” is in English and is not phonologically integrated into Adāηme. Thus, this example clearly rejects the prediction of the PF Disjunction Theorem.

### 3.2.3.3.1. Evaluation

The main argument of Macswan's (2000; 2005) PF Disjunction Theorem is that switching of phonological systems between a stem and an affix is prohibited. Most of the morphemes in isolating languages like Mandarin and Southern Min are free morphemes, which can occur independently, the prediction of the PF Disjunction Theorem, similar to Poplack's (1980) free morpheme constraints, were found not applicable to most of the Mandarin/Southern Min CS data. The first aim of the present



study is to answer the question: are there universal constraints on CS? Because of the counter-examples reported in previous literature and the inapplicability of the PF Disjunction Theorem to the Mandarin/Southern Min CS data, Macswan's (2000; 2005) minimalist model was not adopted to analyse the two sets of data in this thesis.

### 3.3. The Matrix Language Frame (MLF) Model

As suggested in Chapter 1, the MLF (Matrix Language Frame) model is one of the most influential theories in the study of CS. This model, proposed by Myers-Scotton (2002a), views codeswitched utterances from a grammatical and morphosyntactic perspective. The research question this thesis attempts to answer is: 'is there a universally applicable CS model?' Since the MLF model is claimed to have universal applicability, one of the aims of this thesis, then, is to test this model with Mandarin/Tsou and Mandarin Southern Min data.

#### 3.3.1. The Main Arguments of the MLF Model

##### 3.3.1.1. The Content-System Morpheme Opposition

One very important aspect of the MLF model is the distinction between **content morphemes** and **system morphemes**. Myers-Scotton (1993/1997) introduced a feature [quantification] to distinguish these two: those having the feature [+quantification (including quantifiers, specifiers, and inflectional morphology)] are system morphemes, while those morphemes which acquire [-quantification] are potential content morphemes. In a later work, Myers-Scotton (2002a, p.15) states that content morphemes are "the main elements conveying semantic and pragmatic aspects of messages", and system morphemes "largely indicate relations between the content morphemes." The two types of morphemes, according to Myers-Scotton, may also be distinguished through the concept of thematic role assignment. That is, those which

assign or receive thematic roles are considered content morphemes while system morphemes are not involved in the process of thematic role assignment. Hence, content morphemes have the feature “[+thematic assignment]”, while system morphemes have the feature “[-thematic assignment]”. Myers-Scotton (1993/1997) attempted to relate the two types of morphemes to different levels of language production, and claimed that the major emphasis of this model is the abstract procedures directed by lemmas in the mental lexicon.

### 3.3.1.2. The Matrix Language-Embedded Language Opposition

The other key notion in the MLF model is the Matrix Language-Embedded Language opposition. According to Myers-Scotton (1993/1997), a matrix language (ML) is the language that plays the dominant role in a given CS utterance, and the embedded language (EL) refers to the other language involved in the same utterance. It is important to note that in the MLF model, the ML and the EL do not contribute equally. That is why the ML is defined as “more dominant”. Myers-Scotton (1993/1997, p.82) argues that “ML sets the morphosyntactic frame for ML+EL constituents<sup>7</sup> (**The Matrix Language Hypothesis**)”. However, how do we define which language in a CS utterance is the ML? To answer this question, she proposes two principles, namely the **morpheme order principle** and the **system morpheme principle**. The definitions of these two principles provided by Myers-Scotton are shown below.

*“The Morpheme Order Principle:*

*In ML + EL constituents consisting of singly-occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic*

---

<sup>7</sup> ML+EL constituent – those showing morphemes from the two or more participating languages (Myers-Scotton 1993, p.6)

*relations) will be that of the ML.*

*The System Morpheme Principle:*

*In ML +EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e. which participate in the sentence's thematic role grid) will come from the ML.”*

(Myers-Scotton 2002a, p.59)

According to the morpheme order principle the ML determines the order of the elements in mixed (ML + EL) constituents. The system morpheme principle specifies a particular type of system morpheme, which must be supplied by the ML namely the outsider late system morpheme, (c.f. section 3.3.2.).

Myers-Scotton (1993/1997) also introduces two important notions, namely bare forms and EL islands. Bare forms are EL language content morphemes, which do not receive inflections or modifying function words from either the ML or the EL language. The notion EL islands refers to mixed constituents, which are well-formed in the EL and show syntactic independence from the ML (Myers-Scotton 2002a). Neither bare forms nor EL islands are completely integrated with the morphosyntax of the ML. Furthermore, it is worth pointing out that in Myers-Scotton's (1993/1997) earlier version of the MLF model, she used “a sentence” as the unit for analysis; whereas, in her later work (Myers-Scotton 2002a, p.8), she uses “a bilingual CP”<sup>8</sup> as the basic unit of analysis. Based on the concept of maximum projection in syntax, Myers-Scotton (2002a) argues for the use of CP as an analysis point because its status

---

<sup>8</sup> According to Myers-Scotton (2002, p.8), a CP refers to “projection of complementizer”, which includes both “dependent and independent clauses”. A bilingual CP refers to “a CP which contains bilingual constituents” (p.56).

is clear, namely a complementizer and an element in specifier position followed by an IP.

### 3.3.2. The 4-M Model

With reference to the content-system morpheme opposition, Myers-Scotton and Jake (2000) further develop three subcategories of system morphemes, namely early system morphemes, and two types of late system morphemes (bridge morphemes and late outsider morphemes). In addition to the feature “[± thematic assignment]”, they propose two other features, namely “[± conceptually activated]” and “[± referring to grammatical information outside of its  $X^{\max}$  ]” for the purpose of distinguishing these four types of morpheme i.e., content morphemes, early system morphemes, bridge morphemes, and outsider late system morphemes. The feature “[± thematic assignment]” refers to whether a morpheme assigns or receives a thematic role. The feature “[± conceptually activated]” indicates whether the use of a morpheme requires the activation of the lemma which directly links to speakers’ intentions. Finally, the feature “[± referring to grammatical information outside of its  $X^{\max}$  ]” shows that whether a morpheme depends on the information outside its immediate maximal project to acquire its form. To illustrate the main arguments of the 4-M model, consider my examples (25) and (26).

(25) Naomi looked for a job.

(26) He likes the electronic products of Japan.

In examples (25) and (26), *Naomi*, *look*, *job*, *he*, *like* and *electronic product* are content morphemes because they either assign or receive thematic roles. Furthermore, as mentioned above, content morphemes carry the semantic or pragmatic messages whose lemmas are directly related to a speaker's intention. Therefore, a content morpheme must be conceptually activated. Hence, Myers-Scotton (2002a) argues that a content morpheme has the features “[+thematic assignment]” and “[+conceptually activated]”. All nouns and verbs are considered to be content morphemes.

In contrast with content morphemes, early system morphemes have the features “[−thematic assignment]” and “[+conceptually activated]”. In other words, they do not receive or assign thematic roles, but are still related to the speaker's intention. For instance, the determiners “the” and “a”, the preposition “for” after the verb “looked”, the past tense morpheme “-ed” in (25) as well as the plural marker “-s” in (26) are all examples of early system morphemes. This is because they are all indirectly elected by their heads, i.e., content morphemes. For example, the plural marker “-s” acquires its form depending on its head, namely the content morpheme “product”. Myers-Scotton (2000, p.1055) argues that the use of an early system morpheme conveys the “semantic and pragmatic meanings that satisfy the speakers' intention”, and therefore this type of morpheme has the feature “[+ conceptually activated]”. Typical examples of early system morphemes are determiners and plural markers.

Early system morphemes and late system morphemes are distinguished by the feature “[−conceptually activated]”. Unlike early system morphemes, late system morphemes do not require the activation of the lemmas that directly link to speaker's intentions. The “of” in example (26) is an example of a bridge morpheme, whose main function is to “unite morphemes into larger constituents” (Myers-Scotton 2000, p.4). As its

name suggests a bridge morpheme, like “of”, links up the hierarchical relationship between the content morphemes “electronic product” and “Japan”. Possessive marker “’s” and “of” are often considered to be bridge morphemes.

The final type of morpheme is the outsider late system morpheme. Morphemes of this type “depend for their form on information outside their immediate maximal projection” (p.75). A distinction is made between a bridge morpheme and a late outsider system morpheme, again, according to what value of the feature “[± referring to grammatical information outside of its  $X^{\max}$  ]” they have. A late outsider system morpheme needs to acquire its form by checking the information outside its maximal projection, so it has the feature “[+ referring to grammatical information outside of its  $X^{\max}$  ]”. A bridge morpheme, however, does not require such a checking procedure to acquire its form, so it has the feature “[- referring to grammatical information outside of its  $X^{\max}$  ]”. According to Myers-Scotton (2002a), examples of late outsider system morphemes are subject-verb agreement affixes and case affixes. An example of a late outsider system morpheme would be the “-s” which indicates the subject-agreement in example (26). Although “-s” is attached directly to the end of the content morpheme ‘like’, it has to check information outside its immediate maximal projection level (in this case a VP) for its suitable form. The subject “he” in (26) is a 3<sup>rd</sup> person singular pronoun, hence, “-s” must occur.

In short, the 4-M model can be treated as an augmented model of the MLF model. It more precisely clarifies the relationships between different kinds of morphemes and language production processes on the basis of content-system morpheme opposition. It illustrates how different stages of the language production process may select the

required morpheme types and therefore reveals how mental lexicons relate to speech production at the surface level. According to the system morpheme principle, the ML provides the system morphemes, which contain grammatical relations outside their head constituent. The 4-M model makes explicit the fact that these system morphemes are late outsider system morphemes (e.g., subject-verb agreement) because they have to seek their forms outside their immediate maximum projection level. Thus, the identification of the late outsider system morphemes is crucial in identifying the ML of a given bilingual clause.

### 3.4. The Typological Approach

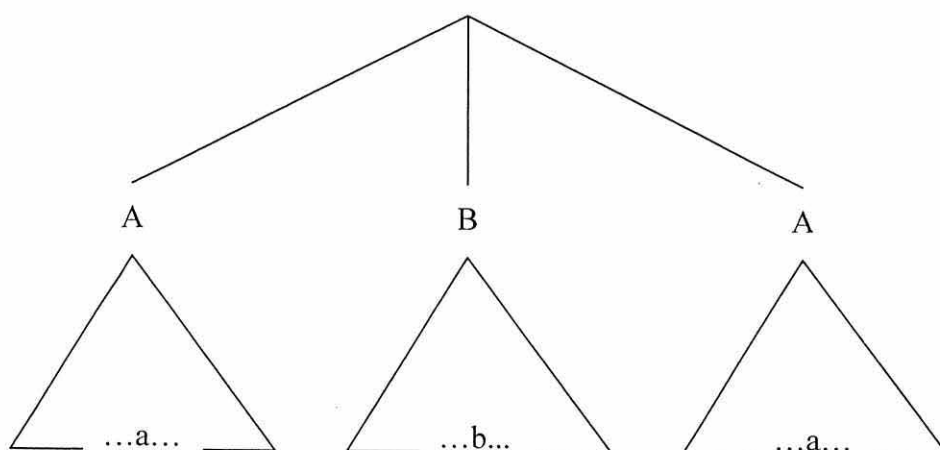
Muysken (2000) adopts both Poplack's (1980) and Myers-Scotton's (1993/1997; 1995) models as well as Labov's (1972) work on standard/dialect contact, and develops a different approach to the study of CS, namely the typological approach. This approach categorizes all the CS utterances into three major patterns: **insertion**, **alternation** and **congruent lexicalization**. It must be noted that Muysken (2000) uses the term "code-mixing" instead of "code-switching". He argues that "code-switching" seems to refer specifically to the second of the patterns he proposes, namely alternation (the alternation between different language structures). Nonetheless, the term "codeswitching (CS)" will still be used when discussing Muysken's model in the present study.

#### 3.4.1. Insertion

According to Muysken (2000), insertion refers to an intra-sentential CS utterance in which the lexical items or constituents from one language are inserted into the structure of the other language. Insertion CS appears to assume a notion from Myers-Scotton's (1993/1997; 2002a) MLF model, namely that there is a matrix

language, which supplies the basic grammatical structure in a codeswitched utterance. The basic structure of insertion is illustrated in (27). In (27), “A” and “B” refer to the two participating languages in the CS, while “a” and “b” are the linguistic elements in these two languages. The linguistic elements in language B are switches, which are inserted into the grammatical frame supplied by the ML, i.e., language A.

(27) **Insertion:**



(Muysken 2000, p.7)

The Swahili/English example in (28) from Myers-Scotton’s (1993/1997, p.86), is used to illustrate insertion CS by Deuchar, Muysken, and Wang (forthcoming, p.3) and is a clear case of insertion. The English word “plate” is inserted into the grammatical structure supplied by Swahili.

(28) A-na-ku-l-a                      *plate* m-bili              z-a              murrām  
 3s-PRES-INFIN-eat-INDIC      plate CL 10-two      CL 10-of      maize  
 ‘He eats two plates of maize’

(PRES = present tense; INFIN = infinitive; INDIC = indicative mood; CL =class)



(Deuchar, Muysken, and Wang; forthcoming, p.3; citing Swahili/English; Myers-Scotton 1993/1997, p.86)

Notwithstanding the MLF model, Muysken (2000, p.68) points out that “the notion of ML is essentially an empirical one ...rather than a theoretical prime”, and, therefore, he argues that there is no single criterion which can be used to clearly define the ML. Instead, he proposes several diagnostic criteria, which help to identify the ML. For example, he argues that most cases of insertion CS are the insertion of *single constituents* (e.g., a single word or a phrase). To illustrate this point, consider Muysken’s example in (29).

(29) Ni-ka-wash all the clothes VP.

1s-consec<sup>9</sup>-wash all the clothes

‘I washed all the clothes.’

(Muysken 2000, p.62; citing Swahili/English; Myers-Scotton 1993, p.80)

In (29), the italic and underlined English words form a full constituent (i.e., a VP). Hence, he argues that the VP in English is a switch, which is inserted into the basic grammatical structure supplied by the ML, namely Swahili. Moreover, Muysken argues that the codeswitched utterance often presents a *nested a... b... a...* structure. “a” refers to the linguistic elements provided by the ML, while “b” is that from the EL. This kind of structure indicates that the elements before and after the inserted switches are grammatically linked. Consider his Quechua/Spanish CS example in (30).

---

<sup>9</sup> consec= consecutive tense

(30) *Chay-ta las dos de la noche-ta chaya-mu-yku.*

that-AC<sup>10</sup> the two of the night-AC arrive-CIS<sup>11</sup>-1<sup>st</sup> pl.

‘There at two in the morning we arrive.’

(Muysken 2000, p.63)

Muysken argues that in (30) although the word *chay-ta* ‘there’ and *chaya-mu-yku* ‘we arrive’ are separated by the inserted elements *las dos de la noche-ta* ‘the two of the night’, they are still grammatically related. Another feature of insertion CS he suggests is that the switches are usually *content words* (e.g., nouns, adjectives, and verbs) rather than function words (e.g., prepositions). Moreover, many cases of insertion CS are *selected elements*, namely objects or complements, rather than adjuncts. The last feature Muysken proposes for **insertion** is that the switched elements of the insertion CS are often *morphologically integrated*, which is illustrated by his Quechua/Spanish CS example in (31).

(31) *Desmaya-chi-pu-ni nuqa-pis.*

Faint-CAU-BEN-1sg I-also

‘I also let (him) faint.’

(Musyeken 2000, p.64)

To sum up, Muysken (2000, p.64) argues that insertion switches are often “single, nested, often selected, often morphologically integrated constituents, often content words.”

---

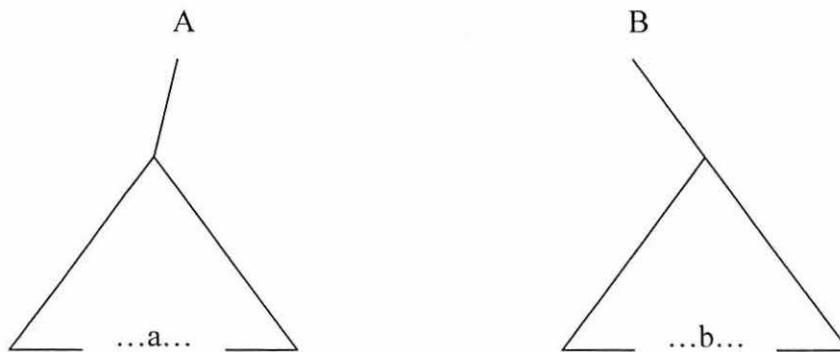
<sup>10</sup> AC refers to accusative case.

<sup>11</sup> CIS refers to cislocative (near or toward speaker).

### 3.4.2. Alternation

Alternation CS refers to the alternation of the grammatical structures from one language to the other in a given codeswitched utterance; this can be illustrated by the structure in (32).

(32) **Alternation:**



(Muysken 2000, p.7)

Muysken (2000, p.4) argues that alternation is associated with Poplack's (1980) model (c.f. section 3.2.3.1.) in terms of "the compatibility or equivalence of the languages involved at the switch point." Hence, in the analysis of alternation CS, the linear word order equivalence is crucial because it implies the introduction of the grammatical structure from the other languages in a single CS utterance. Muysken (2000) also suggests several diagnostic features of alternation CS. The following discussion will briefly review some of these features.

First, he argues that switches of *several constituents* in sequence may be highly likely, for two different grammatical systems are involved in a single utterance. To illustrate this, consider the example in (33) in which the italic switch "*daan vinger hier*" contains more than one constituent.

(33) Je dois glisser *daan vinger hier*.

‘I have to insert/my finger here.’

(Muysken 2000, p. 96; citing Dutch/French; Treffers-Daller 1994, p.213)

According to Muysken, alternation CS often has a *non-nested a...b...a...* structure. Contrary to the *nested a...b...a...* structure of the insertion CS, the definition of *non-nested a...b...a...* features is that the elements which proceed or follow the switched elements are not structurally related. To illustrate this, consider Treffers-Daller’s French/Dutch CS data (1994, p.204), as shown in (34).

(34) Bij mijn broer *y a un ascenseur* en ales.

At my brother’s place, /there is an elevator /and everything.

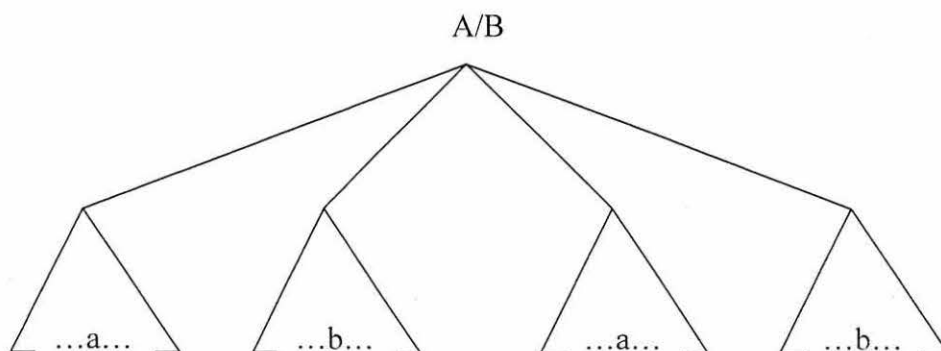
In example (34), the fragments “Bij mijn broer” and “en ales” are not syntactically related and therefore present a *non-nested a...b...a...* structure. He argues that the longer and the more complex a switched fragment is, the more likely that it is a case of alternation. Thus, *the length and complexity* of a switched element can be helpful in identifying whether it is an alternation CS or not. Furthermore, the switch point is also an important criterion to identify the CS pattern. Muysken argues that if a switch occurs in a *peripheral position* (e.g., adverbials), it is likely that the switch is alternation. Other possible diagnostic features he proposes to identify alternation CS include *discourse markers, particles, tags, flagging, dummy insertion* (semantically empty elements), and *self-corrections*.

### 3.4.3. Congruent Lexicalization

According to Muysken (2000), congruent lexicalization refers to the different

languages/language varieties involved in the CS utterance, either entirely or partially, sharing a grammatical structure, while the vocabulary come from two or more languages. The best example would be a CS between a standard language and its dialects. To the sociolinguists like Labov (1972), the term CS should be distinguished from standard language/dialect variation or style shifting. Labov argues that CS is the alternation between two different language systems in a single discourse, while standard language/dialect variation or style shifting refers to free variation within the same system. Muysken (2000, p.124-125) however argues that Labov takes “a paradigmatic approach, and focuses on different realization of the switched elements”, while the study of CS takes “a syntagmatic approach and focuses on the sequential relation of the elements in the sentence”. In other words, both approaches look at the same thing but from different perspectives. Therefore, Muysken argues that “the phenomenon of style shifting can be seen as one subtype of code-mixing, namely congruent lexicalization” (p.123). The basic structure of congruent lexicalization is illustrated by (35).

**(35) Congruent Lexicalization**



(Muysken 2000, p.8)

Muysken proposes two possible reasons for the occurrence of congruent lexicalization.

First, there is an abundance of homophnouns words or diamorphs in the participating languages, which could trigger CS. Second, since the participating languages have similar grammatical structures (both categorical and linear), the constraints on the switch points are not as strict as in the previous two patterns of CS (i.e., insertion and alternation). Therefore, switches may occur in any position. To illustrate this, consider the Ottersum dialect (*italic*)/standard Dutch example in (36).

(36) *ja maar bij ouwe mensen komt dat gauwer tot stilstand als bij jonge mense wa*

“yes but with /older people/ comes that/more quickly to a halt than / with younger people eh”

(Muysken 2000, p.130, citing Giesbers 1989, p.147)

As the languages/language varieties involved in CS have similar grammatical structures and vocabularies (especially so in the case of standard language and dialects), the presence of the diagnostic features, such as *homophonous diamorphs*, *triggering*<sup>12</sup>, *mixed collocation* and *idioms*, *back-and-forth switches*, *switches of functional elements*, *selected elements*, or *multi-constituents* is then expected. Furthermore, it is often the cases that the two languages/language varieties also have similar morphosyntactic structures. Hence, *morphological integration* is also a possible feature of congruent lexicalization.

#### 3.4.4. CS Patterns and Extra-linguistic Factors

Muysken (2000) argues that the three CS patterns can also be linked to different extra-linguistic factors. The occurrence of insertion is favoured in colonial settings or recent migrant communities in which one language has higher socioeconomic prestige

---

<sup>12</sup> Triggering: the use of a single word in one language may trigger a continuous use of that language.

than other languages. Alternation usually occurs in stable bilingual communities in which two or more languages share more or less equal status. Congruent lexicalization, as mentioned above, is most likely to occur when the relationship of the participating languages are a standard language and its dialect.

### 3.5. A Cross-disciplinary Approach to CS

All the approaches introduced above are derived from the theories of socio-linguistics or two linguistic disciplines - socio-pragmatics and morphosyntax. In his research into the bilingual or multilingual speech of European and Asian immigrants in Australia, which includes CS utterances between European languages/English and Asian languages/English, Clyne (2003) argues that current CS research seems to focus too much on lexical and syntactic perspectives, and ignores the importance of the language contact phenomenon, which occurs in other linguistic levels (e.g., semantic, and phonological). He introduces the term **transference** to refer to the language contact phenomenon, which is otherwise known as CS. Clyne (2003, p.76) argues that “a transfer is an instance of transference, where the form, feature or construction has been taken over by the speaker from another language, whatever the motives or explanation for this”. He goes on to argue that transference as a concept has the advantage of covering a much wider range of aspects, while traditionally the term CS only refers to the switches of linguistic elements at morphosyntactic level. The types of transference he provides include: lexical transference, multiple transference, morphemic transference (transference of bound morphemes), syntactic transference, phonological transference, semantic transference, and pragmatic transference (e.g., the use of discourse markers like ‘well’).

Syntactic transference simply refers to the transfer of syntactic rules from one

language to another. For instance, the Mandarin Chinese Det + N construction is replaced by Tsou's N + Det order. Phonological transference is the addition or deletion of phonemic structure because of the influence of the other language. The example Clyne (2003, p.78) provides is that “/ts/ [is] replaced by /s/ in third-generation German-English bilinguals”. Semantic transference occurs when the meaning of the words in one language morphemically and semantically correspond to the words in other languages and is illustrated by my example in (37):

(37) ta de *gakko* hao da. (invented Mandarin/Japanese CS)  
3s poss school very big  
'His school is very big.'

Following Clyne's argument that the Japanese word *gakko* 'school' in example (37) is identical to *xue-xiao* 'school' in Mandarin in terms of its characters and meaning, it is therefore switched.

It should be noted that the notions of “lexical transference (transference of a single lexical items)” and “multiple transference (transference of two or three consecutive words which may form a constituent)” correlate to Muysken's (2000) insertion CS (Clyne 2003, pp.73-74). Further, Clyne (2003, p.80) introduces another term – **transversion** which refers to “a crossing over from one language to another rather than a transference of an item, feature or construction”. In other words, a transversion includes “both intra- and interclausal (CP or sentential switching)”, which covers Muysken's (2000) alternation CS and congruent lexicalization (Clyne 2003, p.76).

In his later discussion, Clyne (2003, p.162) proposes a notion – “facilitation” to



replace a more general term “constraint”, and provides three facilitation principles to predict the places that a switched element is likely to occur. Now, consider these three principles.

“Facilitation Principle 1:

Lexical items that can be identified as being part of more than one language for the speaker or for some section of, or the entire speech community, may facilitate a transversion from one language to another (including the triggering words e.g. lexical transfer, bilingual homophones, and proper nouns).

Facilitation Principle 2:

Lexical items in a tonal language whose tone is identified with the pitch and stress of the non-tonal language in contact are liable to facilitate (though not necessarily cause) transversion.

Facilitation Principle 3:

If syntactic rules overlap between the languages or previously divergent syntactic rules converge in the individual grammars of the speakers, switching is facilitated.”

(Clyne 2003, pp.162-179)

Principle 1 and 3 are more or less generally agreed ideas in the CS literature, but they still have the problem of a lack of a basic analysis unit. The example Clyne provides to illustrate principle 2 is also problematic. He argues that “in Mandarin-English language contact, it is falling tones, fourth, half-third and neutral that facilitate transversion” (p.176). Such an argument is proposed based on Zheng’s (1997)

research into Mandarin/English CS utterances produced by second and young first-generation Mandarin-English bilinguals in Melbourne. Zheng claims that “about 96.49% of switches came after such a tone, corresponding to English pitch and stress” (p.176). However, I argue that it is problematic to draw a link between “stress” and “pitch” in English and “tone” in Chinese. In English, many words can be differentiated by “stress.” Consider Crystal’s (1997, p.364) examples in (38).

- (38) a. An increase in pay is needed.  
b. I’m going to increase his pay.

In (38a), the stress of the word “increase” is in the beginning, and it is pronounced as “/ɪŋkri:s/”. In (38b), the stress of “increase” is in the middle, and it is pronounced as “/ɪŋ'kri:s/”. The examples in (38) show that in English “stress” may change the word class (in this case changing from a noun to a verb), but the basic meaning “to make something greater” remains unchanged. However, in Chinese, tone has completely different functions. Consider my examples in (39) and (40).

(39) mā (first tone) = mother (noun); to wipe (verb)

má (second tone) = numb (adjective)

mǎ (third tone) = horse (noun)

mà (fourth tone) = to condemn (verb)

(40) a. ni    **mā**    hao    piao-liang.

you mother very beautiful

‘Your mother is very beautiful.’

b. ni    mǎ    hao    piao-liang.  
you horse    very beautiful  
'Your horse is very beautiful.'

As (39) and (40) show, the change of tone can make a significant difference in terms of meaning and word class. Hence, it is problematic to treat “stress” in English as “a correspondence” to “tone” in Chinese. A more detailed discussion of Zheng’s (1997) work will be provided in section 3.7.2.

In summary, Clyne’s model of CS phenomena is cross-disciplinary, and covers almost every linguistic aspect. However, unlike Myers-Scotton’s (2002a) MLF model which specifies a CP as the unit of analysis, the analysis point in Clyne’s (2003) approach is rather vague. By adopting a morphosyntactic theory, such as the MLF model, one only needs to look at the morphosyntactic frame and the interaction of the grammatical systems involved in a given CS utterance and predict the position where a morphosyntactic unit is likely to be switched. The range covered by Clyne’s model, however, is too broad. Hence, it is difficult to focus on any particular linguistic aspect to answer questions such as: what linguistic elements (e.g., syntactic, semantic, phonological or morphological etc) will be switched and where will such switched elements occur? Thus, Clyne’s model was abandoned by the present study.

### 3.6. Discussion of Two Morphosyntactic Approaches

It has been stated that both Myers-Scotton (the MLF model and its supplementary models) and Muysken (the Typology approach) tackle the CS issue from a morphosyntactic perspective. Although some of their main arguments overlap, there remain many differences. In this section, the similarities and differences of these two

models will be discussed. Then, an evaluation by other linguists of the two approaches will be presented.

### 3.6.1. A Contrastive Analysis of the Two Approaches

Among the several similarities and differences between Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach, the most fundamental difference is the universal applicability of these two models. Myers-Scotton (2006) argues that no matter what languages are involved in a given bilingual clause, the two grammatical constraints, namely the morpheme order principle and the system morpheme principles, of the MLF model can be applied successfully. In other words, according to her argument there is only one language (i.e., the ML) which contributes more by supplying the morphosyntactic structure of a given bilingual CP, while the other participating language plays a less important role by merely providing certain types of morphemes to fit in the grammatical structure set by the ML. She argues that the two grammatical constraints can always unambiguously identify the ML of a given bilingual clause in which any language pairs participate. Muysken (2000), however, clearly states that there is no single model, which can cover all CS phenomena. He argues that Myers-Scotton's (2002a) MLF model only works for one pattern of CS, namely insertion. To account for all CS phenomena, one must adopt a more comprehensive perspective. Hence, he proposes the other two CS patterns, namely alternation and congruent lexicalization.

The second difference between these two models is that the MLF model only looks at intra-sentential (or intra-clausal) switches, whereas the typological approach not only examines intra-sentential CS, but also inter-sentential CS and the switches which occur between different turns. Third, Muysken's insertion is derived from the basic

argument of the MLF model, namely there is a ML which provides the basic grammatical structure of a given codeswitched utterance. Nevertheless, the way he identifies the ML is different from Myers-Scotton's (2002a). Myers-Scotton proposes two important criteria to identify the ML, namely the morpheme order principle and the system morpheme principle. Muysken (2000), however, argues that no single principle or grammatical constraint can be used to identify the ML. Instead, he introduces several diagnostic features for the identification of the ML.

Fourth, Myers-Scotton (1993/1997; 2002a) argues that the languages involved in an intra-sentential CS are not treated equally. There is always a language which contributes more in a CS utterance and that is the ML. Other languages which participate in the given CS utterance are the EL. The ML sets the morphosyntactic frame of the ML+EL constituents, while the EL supplies content morphemes, early system morphemes, and bridge morphemes to be inserted into that frame. Although this scenario holds for Muysken's (2000) insertion, it is not applicable to alternation and congruent lexicalization. Alternation involves two or more independent morphosyntactic systems, which may participate equally in a codeswitched CS utterance. Hence, it is hard to define which language involved in the CS is dominant. In congruent lexicalization, the involved languages may have very similar morphosyntactic systems. Therefore, it is also difficult to decide which language supplies the basic grammatical frame.

### 3.6.2. Evaluation of the Two Approaches

#### 3.6.2.1. The MLF Model

The MLF model and its augmented model (i.e., the 4-M model) have been tested with CS data involving various language pairs. For instance, by examining the variation in

gender assignment in German/English CS data, Fuller and Lehnert (2000) point out that their findings support the MLF model, for it accurately predicts that the grammatical gender assignment is provided by the ML (i.e., German). However, Boussofara-Omar (2003) proposes some potential counter-examples to the MLF model and the other two augmented models with reference to her Tunisian Arabic (TA) and Fushaa (F) CS data. Consider her example in (41).

(41) ma-ʔa-ðunnu-ʃ                      kɛɛnu

NEG-1SG IMP- think-NEG      were

TA – F- F- TA                      TA

‘I do not think they were’.

(NEG = negative; SG = singular; IMP = imperfect)

(Tunisian Arabic/ Fushaa; Boussofara-Omar 2003, p. 39)

She claims that in example (41) both TA and F provide system morphemes within the same CP. The negation marker “ma-.... -ʃ” in TA and the first person singular imperfective tense marker “ʔa-” in F are system morphemes and are assigned simultaneously to the verb (a content morpheme) *ðanna* ‘think’. Hence, she argues that this example violates the system morpheme principle.

Boussofara-Omar’s discussion seems to indicate a misinterpretation of the system morpheme principle. Myers-Scotton (2002a, p.87) argues that “the system morpheme principle does not state that all the system morphemes must come from the matrix language.” In fact, what this principle states, is that “all system morphemes which have grammatical relations external to their head constituent will come from the ML”



Myers-Scotton's (2004, p.89) response to this counter-example is that "the MLF model was formulated to cover codeswitching between language varieties that are separate languages (i.e., not mutually intelligible varieties, such as dialects)". I argue that Myers-Scotton's response is problematic in two ways. First, the definition of a dialect is a controversial issue. In fact, many languages are categorised as "dialects" of a certain language because of socio-political rather than linguistic factors. It is therefore problematic to assume that all dialects are linguistically similar. Second, even if two languages are typologically similar, the speakers of these two languages are not necessarily mutually intelligible. One of the language pairs this thesis aims to test, namely Mandarin and Southern Min, is a good example. Southern Min is often categorised as a dialect of Chinese, and it shares most of its morphosyntactic structure with Mandarin. However, a monolingual Mandarin speaker and a monolingual Southern Min speaker are not intelligible to each other. This is because Mandarin and Southern Min differ significantly in terms of their lexicons and phonology.

Macswan (2005) also criticises the MLF model. Consider his example in (43).

(43) ¿Funciona *the computer* de tu hermano en la oficina?

Function-3Ss of your brother in the office

'Does your brother's computer in the office work?'

(Spanish/English; Macswan 2005, p.10; citing Macswan 1994)

In (43), the ML is Spanish and the EL is English. Macswan argues that the EL also supplies a system morpheme (i.e., the definite article "the"), and therefore this example should be regarded as a violation of the system morpheme principle. As in Boussofara-Omar's (2003) discussion, Macswan's analysis shows a misinterpretation



of the MLF model. The MLF model does not state that all system morphemes must come from the ML. According to the 4-M model (c.f. section 3.3.2), it is the late outsider system morpheme, which has to be provided by the ML. The definite article “the” is an early system morpheme, which does not necessarily come from the ML. Hence, (43) does not appear to violate the system morpheme principle.

#### 3.6.2.2. The Typological Approach

Muysken’s (2000) typological approach has been criticised by several linguists. For instance, Myers-Scotton (2002b) argues that in the beginning of his book, Muysken (2000) clearly states that the distinction between CS and lexical borrowing has to be made. However, in the later discussion of the data he provides, Muysken defines lexical borrowing as a special type of insertion. At another point, when referring to same data in explaining lexical borrowing, he uses insertion. Thus, Myers-Scotton (2002b) argues that the status of whether lexical borrowing is CS or not remains obscure in Muysken’s model.

Poplack and Walker (2003) point out several problems with Muysken’s model. One of the most important points they raise is that some diagnostic features Muysken proposes are shared by the three CS patterns, namely insertion, alternation, and congruent lexicalization. They argue that, when analysing empirical data, researchers may have trouble in identifying the actual pattern of a switch. For instance, the feature *morphological integration* would not distinguish the pattern of a given switch as insertion or congruent lexicalization, for both CS patterns have such a feature.

According to Muysken’s (2000) model, the identifying of the CS pattern of a given switch does not depend on a single diagnostic feature. In fact, twenty-seven features

in total have to be examined for the identification of the CS pattern. Hence, in contrast with Poplack and Walker's (2003) view, I do not consider the fact that some diagnostic features are shared by different CS patterns to be a theoretical problem of Muysken's (2000) model.

### 3.7. Earlier CS Research Relevant to Chinese

#### 3.7.1. An Integrated Approach to Language Choice Patterns and Chinese/English CS

Li Wei (1994) adopted an integrated approach to examine language choice patterns and CS phenomena of the Chinese community in Newcastle Upon Tyne, UK. As mentioned in section 3.2.1.1, a pure sociolinguistic approach to CS is sometimes criticised for paying too much attention to macro-linguistic factors that may influence CS phenomena and ignore the effects that come from a more local level (e.g., conversational structure or speakers' preference and competence). In contrast, a pure CA approach is sometimes criticised for ignoring the influences from a wider political and socio-economic context upon CS. To avoid those problems generated by a pure sociolinguistic or CA approach to CS, Li Wei adopted both approaches and conducted a two-stage analysis to look at language choice patterns and CS phenomena both from a community level and a conversational interaction level.

At the first stage of analysis, Li Wei observed the language choice patterns of the 58 subjects from Tyneside Chinese families with reference to the sociolinguistic variables, such as age, gender etc. The results of his observation suggested that there was "an age-related language shift from Chinese monolingualism to English-dominant bilingualism" (Li Wei 1994, p.179). In other words, the language choice pattern of the parent and grandparent generations was mainly Chinese-dominant, while that of the child generation was English-dominant.

He argued that, in addition to the sociolinguistic factor of age, the social networks of the selected subjects also provided important clues to explain the process of language shift within the Chinese community. Three types of social networks, namely exchange, interactive, and passive, with regard to ethnic and peer-groups were examined. According to Li Wei (1994, p.118-119) exchange networks are “strong ties (e.g., kin and close friends)”; interactive networks are “weak ties (e.g., a shop owner and his/her customers)”; passive ties refer to “relatives or friends who are physically distant from the subjects”. He found that the grandparent generation had strong exchange networks with other members in the Chinese community, and used Chinese as the major medium of communication. The parent generation had some non-Chinese ties, but they still acquired limited English. In contrast to the previous two generations, the child generation had English-dominant language choice pattern, and most of their social networks are non-Chinese and peer-group-based.

At the second stage of analysis, Li Wei adopted a turn-by-turn CA approach to examine the English/Chinese CS data collected from his fieldworks. The results of his analysis suggested that the parent and grandparent generations used Chinese as the major medium of communication. Switches to English occasionally occurred in order to achieve some communicative purposes, such as repairing or turn-taking. The child generation used English most often when talking to people of their own generation. Li Wei particular pointed out that “they tended to use English to contextualise dispreferred responses in conversations with parents and grandparents” (p.180).

Li Wei’s integrated approach is a very promising model. It avoids the traditional problems that either a pure sociolinguistic approach or CA approach has, and

examines the language choice patterns and CS phenomena in a Chinese community with a more comprehensive scope. Although the focuses of his model are entirely different from a structural approach to CS which this thesis adopts, it provides a perfect example of showing how Chinese in the context of language contact has been studied. Thus, it is necessary to include Li Wei's approach in this chapter.

### 3.7.2. Tonal Aspects of Chinese-English CS

In her study of Chinese immigrants' children in Australia, Zheng (1997) claims that tone in Chinese is a significant factor for the occurrence of Mandarin Chinese/English CS utterances. The children were aged from six to ten years old. The total thirty subjects were divided into six groups with reference to their sex and age: 9 –10 year old, 7 – 8 year old, and 5 –6 year old. In Mandarin Chinese, there are four basic tones. The definitions of these tones provided by Zheng (1997, pp.54-55) are: (1) *First tone*: a high pitch sound (marked by “ˊ”). (2) *Second tone*: a rising sound from middle pitch up to high pitch (marked by “ˊ”). (3) *Third tone*: the curved one from middle-low-pitch down to low-pitch and up to middle-high-pitch (marked by “ˇ”). (4) *Fourth tone*: the falling one from high-pitch down to low-pitch (marked by “ˋ”). Furthermore, there is also a neutral tone, which is applied to syllables which are unstressed or take a weaker tone (pp. 54-55). Consider her example in (44).

(44) Mandarin    English

ai            I

ài            ai

love        I

(Zheng 1997, p.54)

Zheng points out that the pronunciation of the Mandarin word *ai* 'love', as illustrated by (44), is roughly the same as English "I". Hence, she argues that, because of this phonological similarity, the fourth falling tone facilitates a CS from Mandarin to English. From this observation, she claims that the fourth falling tone "plays a very important role in the children's CS to English" because it corresponds to what she calls "normal intonation in English" (p.54). She goes on to argue that in her data the falling tones (i.e., the fourth, the third and the neutral tones) significantly facilitate the children's switching between Mandarin and English. Consider her example in (45).

(45) qì-ér      de      yán-sè      shì      *grey*.

ʈʂí      də      íansè      ʃì

penguin    poss    color    cop

'The color of penguin is grey.'

(Mandarin/English; Zheng 1997, p. 55)

In (45), the Mandarin copular "shì" before the switched English word "grey" is a falling tone, namely a fourth tone. Again, Zheng claims that such a tone facilitates Mandarin/English CS. She concluded that 96.54% of her Mandarin/English CS data have the falling tones (i.e., the fourth falling tone, the half third falling tone, and the neutral falling tone, and the neutral falling tone) before the switches to English.

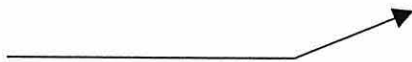
Zheng's argument provides an interesting aspect of explaining why people switch codes (i.e., from a purely acoustic perspective), however there are problems with the

approach, First, her claim, that the Mandarin fourth tone corresponds to English normal intonation, is problematic, for tone in Chinese is a different notion from intonation in English. According to Crystal (1997, p.202), intonation in English is usually used for “(a) marking the contrast between some grammatical structures: e.g., statement vs. question (b) expressing personal attitude: e.g., puzzlement, anger etc”. Hence, different intonation in English only turns a statement into a question (or vice versa), or expresses the speaker’s emotion and attitude, while the meaning of each word remains the same. Now, consider my examples in (46).

(46) a. We will have an exam tomorrow (falling intonation).



b. We will have an exam tomorrow (rising intonation)?



(46a) and (46b) show that the tone either rises or falls at the last word “tomorrow”. However, the meaning of this single word “tomorrow” does not change. Consider the Mandarin examples in (47).

(47) a. **āi**: sad

b. **ái**: cancer

c. **ǎi**: short

d. **ài**: love

Unlike the English examples in (46), the Mandarin examples in (47) show that if the

tone of a Chinese word is changed, its meaning is also changed. It can be seen that the meaning of the basic morpheme “ai” changes if different tones are added to it. Hence, I argue that tone in Mandarin and intonation in English are completely different, and it is to be doubted if these two can be treated as a similar notion which may trigger codeswitched utterances.

From a purely phonological point of view, Zheng’s claim is also problematic. Although her examples in (44), namely the Mandarin word *ài* ‘love’ and the English word “I”, do “sound” similar and may facilitate children’s switching between English and Chinese, the other example she provides in (45) is clearly different. In (45), the Mandarin copula *shì* ‘be’ shows clear difference from the English word “grey” in terms of their pronunciation and tones. It is, then, problematic to claim that a CS example like (45) is facilitated by phonological factors. In fact, in spite of Zheng’s claim that falling tones in Chinese facilitate switches to English, it is not uncommon to see an English switch occur after a Chinese word with the rising tone. Consider Wei’s (2001, p.157) example in (48).

(48) ni    nei   piān   *article*   hai   méi   *finish*   a?

you   that   class                    yet   neg                    part-Ques

‘You haven’t finished that article yet?’

(Mandarin/English; Wei 2001, p.157)

In (48), the Chinese negation marker *méi* ‘not’, which occurs before the English switch “finish”, is the second tone, namely a rising tone. Furthermore, the Chinese classifier “piān”, which occurs before the English switch “article”, has the first tone.

According to Zheng's (1997) definition of tones, the first tone is neither treated as a rising tone nor categorised as a falling tone. Thus, Wei's (2001) example in (48) suggests that the falling tones in Chinese do not necessarily facilitate switches to English.

The discussion above demonstrated two points. First, tones in Chinese are linguistically different from intonations in English and therefore should not be treated as similar elements. Second, since the switches to English, as shown in (48), may also occur after the first tone and the second rising tone, the falling tones do not necessarily facilitate CS to English. However, there must be a reason why 96.54% of Zheng's Mandarin/English CS data have the falling tones before the switches to English. I argue that such a result was caused by the subjects' language proficiency in Mandarin. Note that the subjects were aged between six to ten years old and they were Chinese immigrants in Australia, an English speaking country. Hence, their proficiency in Mandarin was still at a developing stage. This argument is supported by the fact that the examples provided by Zheng (1997) were mainly simple sentences in Chinese. Consider her examples in (49) and (50) below.

(49) kàn diàn-shì, *go to bed*.

Watch TV

'(I) watch TV and then go to bed.'

(Zheng 1997, p.55)

(50) (panda) hěn hǎo – de *happy*.

(panda) very well nom.

'The panda is very well and happy.'



(Zheng 1997, p.56)

Almost all of the examples Zheng provided in her study resembled the examples in (49) and (50), which have rather simple sentence structures. In fact, the example in (50) is not a correct sentence. The nominalizer *de* should not occur with reference to Chinese grammar. Hence, this example further supports the argument that the Zheng's subjects were still developing proficiency in Mandarin. In addition to that, there is a rather high degree of the repetition of some linguistic elements in the twenty-four examples provided in Zheng's study. Consider her examples in (51) and (52).

(51) zài city. city hěn rè<sup>4</sup>nào.

in very bustling

(Zheng 1997, p. 58)

(52) xian-zài ta zài bicycle shàng ride.

Now he on on

'Now, he is riding on a bicycle.'

The preposition *zai* 'in/on' occurs in both (51) and (52) as well as in other examples she provided. Other linguistic elements such as the word *dian-shi* 'television', the classifier *ge*, the adverb *hen* 'very', the postposition *shang* 'on' etc. occur at least twice or more in only twenty-four examples in Zheng's study. This, again, might be caused by the subjects' limited competence in Mandarin. It is not difficult to imagine that there might be a higher degree of repetition of these or other linguistic elements in Zheng's Mandarin/English CS corpus. It is clear that the frequency of the switches to English before the falling tones was overvalued, for the appearance of the same

linguistic elements were repeatedly analysed. Hence, I argue that Zheng's tonal approach to examine the Mandarin/English CS phenomenon is problematic.

As the major differences between Mandarin and Southern Min largely remain at the phonological level, it was assumed that Zheng's phonological model to CS could have been a useful approach to analyse the Mandarin/Southern Min data. Nevertheless, as discussed above, many theoretical and methodological problems of Zheng's model were found, and therefore it was abandoned by the present study.

### 3.7.3. Morphosyntactic Approach to Chinese-English CS

By examining Chinese/English CS data collected in Singapore, Kamwangamalu and Lee (1991) aimed to answer two questions: (1) whether there is a matrix language in any given bilingual sentence? (2) If the answer is "yes", how is the matrix language to be identified? Kamwangamalu and Lee conducted an experiment in which the participants were asked to take an oral test and a written test. Fifty-nine subjects participated in this experiment; they were all fluent speakers of Chinese and English. In the oral test, the subjects were asked to identify whether the ten sentences given were Chinese-based or English-based (the term "matrix language" will be used in the following discussion). Two weeks later, the same subjects completed the written test, which included the same ten sentences as in the oral test. Again, as in the oral test, the subjects were asked to identify the matrix languages of the ten sentences. They were also asked to make comments on how they had made the judgement. The criteria adopted by the subjects to make the judgment could be divided into two major categories, namely lexical and structural. Lexical criteria<sup>13</sup> include lexical

---

<sup>13</sup> Lexical substitution refers to "the use of single lexical items of one language in discourse in the other language, and lexical density means the proportion of lexical items of one language relative to the other in a CS sentence" (Kamwangamalu and Lee 1991, pp.252-254).

substitution, discourse markers, and lexical density; the structural criteria are structural substitution, word order, verbless clauses, and sentence final particles.

Kamwangamalu and Lee (1991) argued that the lexical criteria such as lexical substitution and discourse markers did not work very well in determining the matrix language. For instance, if there is an English lexical substitution in a CS sentence, the matrix language is not necessarily English. To illustrate this, see their example in (53).

(53) ta            de    *teacher*   bu   hui   *explain*.

His/her    poss.                    neg   able

‘His/her teacher is not able to explain.’

(Chinesen/English; Kamwangamalu & Lee 1991, p.253)

In (53), there are two examples of English lexical substitution, namely *teacher* and *explain*. They argued that if the criterion “lexical substitution” works, then the matrix language should be English. However, they found that seventy-five percent of the subjects reported that Mandarin was the matrix language of (53) in the oral test, and sixty-six percent in the written test.

Kamwangamalu and Lee (1991, p.258) then argued that “the structural clues are more crucial in determining the ML (matrix language) of a code-switching sentence.” The structural clues they referred to were word order, verbless clauses, and sentence final particles. Consider the examples in (54) to (56).

(54) *I think you*   mai   khi   ga   ho. (word order)

don’t   go   comp   good

‘I think you had better not go.’

(Chinese/English; Kamwangamalu & Lee 1991, p.256; citing Tan 1998, p. 86)

If we compare the Chinese/English CS data in (54) with its English translation, it is clear that this CS utterance has Chinese word order. Hence, Kamwangamalu and Lee (1991) argued that Chinese is the ML of (54).

(55) *Grace tai blind.* (verbless clause)

too

‘Grace (is) too blind.’

(Chinese/English; Kamwangamalu & Lee 1991, p.257)

It is clear that there is no main verb in (55). Since verbless clauses or sentences are very frequently found in Chinese, Kamwangamalu and Lee then argued that the ML of (55) is also Chinese.

(56) *This is the one wa ga li gong a la.* (sentence final particle)

I and you tell asp-perf part.-impat

‘This is the one I told you about.’

(Chinese/English; Kamwangamalu & Lee 1991, p.258; citing Tay 1984, p.416)

Kamwangamalu and Lee 1991 argued that the presence of the sentence final particle *la* indicates that (56) has Chinese sentence structure. This is because there is no sentence final particle in English. Kamwangamalu and Lee’s findings entirely depended on the subjects’ intuition, which is not a reliable method. However, they

proposed some interesting criteria which could be used to identify the ML. Some of the criteria they suggested will be adopted and re-tested by the Mandarin/Southern Min data (c.f. section 6.2.2.3).

### 3.8. Conclusion

This chapter has reviewed a range of perspectives on CS in the current literature, including sociolinguistic, CA, structural, and cross-disciplinary approaches. Previous studies related to Chinese CS were also discussed. Since one of the major aims of this thesis is to test Myers-Scotton's (2002a) MLF model and Muysken's (2000) approach, a detailed discussion of these two models was provided, including a presentation of their main arguments, a contrastive analysis between the two models, and criticism of the models raised by other linguists.

## Chapter 4: An Introduction to Mandarin, Southern Min, and Tsou

### 4.1. Introduction

The “Chinese language” is a term often used to refer to all varieties of Chinese spoken in different regions in China and which belong to the Sino-Tibetan language family (Ruhlen 1987). Two languages which this thesis will investigate, namely Mandarin and Southern Min, are dialects of Chinese and are typologically very similar. First, they share most of their syntactic structure, both have SVO word order, and differ mainly in their phonology and lexicons. Second, both languages lack inflectional morphology to express grammatical information, such as tense or subject-verb agreement. Chinese languages are often categorized isolating languages, which have comparatively much less morphology if compared to the agglutinating and inflecting languages (Crystal 1997). Monolingual Mandarin and Southern Min examples are provided in (57a) and (57b) to show the syntactic and morphological similarities of these two languages.

(57) a. wo da xiao-hai. (Mandarin)

I hit child

‘I hit a child.’

b. wa pa in-na. (Southern Min)

I hit child

‘I hit a child.’

Comparison of (57a) and (57b) shows that each Mandarin word corresponds to each Southern Min word in terms of its syntactic position. The only differences are in the lexicon and phonology. Moreover, there are no inflectional morphemes in either (57a)

or (57b) to mark tense.

The Tsou language belongs to the Austronesian language family, though all of its speakers are living in a Chinese-dominant region, Taiwan. As an Austronesian language, Tsou has a complex morphological system. Its basic word order, VOS, is different from the SVO order of the two varieties of Chinese mentioned above. This is illustrated by my examples in (58a) and (58b) and Chang's (1998, p.16) example in (58c).

(58) a. Alon zhua shan-zhu. (Mandarin - SVO)

Name caught wild boar

'Alon caught a wild boar'

b. Alion lia juan-di. (Southern Min – SVO)

Name caught wild boar

'Alon caught a wild boar'

c. mo mayo do fuzu ?o Basuya. (Tsou – VOS)

tense2,Agent catch Obl2 wild boar Nom4 Name

'Basuya caught a wild boar.'

(Chang 1998, p.64)

Observe the similarities and differences between the three examples in (58). We can see in (58a) and (58b), the word order of Mandarin and Southern Min are both SVO, while that of Tsou, as shown in (58c), is VOS. Furthermore, (58c) also reveals that case and tense marking are expressed by the inflectional morphology of Tsou, whereas Mandarin and Southern Min do not mark case and tense. Hence, Tsou is

typologically very different from Mandarin and Southern Min.

#### 4.1.1. Aims of this Chapter

One of the research aims of the present study is to test Myers-Scotton's (2002a) MLF model in which word order and a type of inflectional morpheme, outsider system morphemes, are two important criteria in identifying the matrix language (ML) that supplies the morphosyntactic structure of a bilingual CP. Thus, the discussion in this chapter will focus on the introduction of some syntactic (e.g., the basic word order) and morphological features (e.g., inflectional morphology) of Mandarin, Southern Min, and Tsou in order to provide sufficient background knowledge to help in the identification of the MLs of the Mandarin/Southern Min and Mandarin/Tsou CS data.

According to the MLF model and its augmented model the 4-M model (c.f. section 3.3.2.), there are three subcategories of system morphemes, namely early system morphemes, bridge late system morphemes, and outsider late system morphemes. It is the outsider late system morpheme, which is crucial to identifying the ML. However, to my knowledge, there is no current literature, which discusses the question: what are the early system morphemes, bridge late system morphemes, and the outsider late system morphemes (if there are any) in Mandarin, Southern Min and Tsou? Before one can apply the MLF model to the Mandarin/Southern and Mandarin/Tsou CS data, it is important to identify the types of morphemes in these three languages with reference to the 4-M model. Hence, this chapter will have a detailed discussion of the possible candidates for the three types of system morphemes in Mandarin and Southern Min.

Myers-Scotton (personal communication) argues that the aspect markers (e.g., the



perfective aspect marker “le”) are the outsider late system morphemes in Mandarin. However, I consider that argument to be problematic. In this chapter, the functions or usages of different aspect markers in Mandarin and in Southern Min will be illustrated with examples in order to provide necessary information for the discussion, in Chapter 6 of the question: whether the aspect markers in Mandarin or Southern Min may be treated as the outsider late system morphemes?

Myers-Scotton (2002a) argues that the basic unit of analysis of the MLF model is a bilingual clause (i.e., a bilingual CP). In languages such as English, one only needs to check the main verb to identify a CP, for only one verb is allowed in an English CP. Nevertheless, it is not uncommon to see serial verb construction in Mandarin and Southern Min, which means that it is possible to have more than one verb in a CP. In this case, identifying a CP in Mandarin and Southern Min is more complex. Since it is important to make the basic unit of analysis clear, the criterion that will be adopted to identify a Mandarin and a Southern Min CP in this thesis will be discussed.

Furthermore, the syntactic and morphological similarities that Mandarin and Southern Min share, namely most of their surface word order and having limited inflectional morphology, may become potential problems for the application of the MLF model to the Mandarin/Southern Min CS data. Hence, it is likely that the two important principles, i.e., morpheme order principle and system morpheme principle, of the MLF model may not be applicable. In this case, some criteria adopted by other scholars to identify the ML may be helpful in solving the problem.

One of the criteria proposed by Kamwangamalu and Lee (1991) to identify the MLs of their Mandarin/English CS data is to check the sentence final particles in Chinese

(c.f. section 3.7.3.). In other words, if a Chinese sentence final particle is found in a Chinese/English codeswitched utterance, then Chinese is identified as the ML. That criterion could be useful to solve the problem of the inapplicability of the MLF model to the Mandarin/Southern Min CS data. Thus, it is necessary to briefly review the basic functions of some sentence final particles in Mandarin and Southern Min.

Finally, although Mandarin and Southern Min share most of their syntactic structure, there are still cases that show differences in terms of the surface word order. One apparent example is the construction of the so-called A-Not-A question in Mandarin and Southern Min. As will be discussed in Chapter 6, the MLF model can only ambiguously identify the MLs of the bilingual Mandarin/Southern Min clauses with the construction of the A-Not-A question. Hence, it is necessary to introduce briefly the syntactic structure of the A-Not-A question in this chapter.

#### 4.1.2. The Organisation of this Chapter

The organization of this chapter is as follows. Because of the similarities between Mandarin and Southern Min, these two languages will be introduced one after the other to make comparisons convenient. First, the general linguistic features of Mandarin and Southern Min, including basic word order, tense and aspect, will be described. Second, linguistic elements that may be treated as early system morphemes (e.g., classifiers) and bridge late system morphemes (e.g., the possessive marker *de*) will be discussed at some length. More importantly, the functions of the aspect markers in Mandarin and Southern Min, which are possible candidates for outsider late system morphemes, will be explained. Third, the functions of sentence final particles as well the sentence pattern of *A-not-A* questions in Mandarin and Southern Min will be presented.

Finally, in the discussion of Tsou, linguistic features will be introduced which will be helpful to the application of the MLF model to the Mandarin/Tsou CS data, in particular the basic word order and case marking system. Myers-Scotton (2002a) clearly states that case markers are outsider late system morphemes. Thus, this chapter will contain a discussion of the major functions of different case markers in the Tsou language. As there has been limited study of the Tsou language in the current literature, more information on what I call as the tense and thematic role marking system will be introduced in order to supply more general linguistic information about this little known language.

## 4.2. Word Order

### 4.2.1. Mandarin

With respect to the basic word order of Mandarin, there has been considerable debate on this topic. Some linguists claim that the basic word order is SOV (Subject + Object +Verb) while others argue that it is SVO. In the following discussion, literature relevant to the question will be reviewed.

Li and Thompson (1981) argue that it is hard to determine the basic word order of Mandarin because both the SVO and SOV orders are found in modern Mandarin and they serve different semantic functions. They claim that the object noun phrase after the verb is indefinite while the one before the verb is definite. Their examples, namely (59) and (60) are provided below (Li & Thompson 1981, p.21).

(59) wo zai mai shu le.

I asp-dur buy book asp.-perf.

‘I am buying a book.’

- (60) wo shu mai le.  
I book buy asp.-perf.  
‘I have bought the book.’

To dispute Li and Thompson’s claim, Sun and Givon (1985) adopted a quantitative approach and analysed two texts: one written and one oral. The written text was selected from Chapter 3 (about 25 pages) of the Mandarin novel *The First President* (J. Huang et al. 1983). The oral text was transcribed from the narration of three native speakers of Mandarin Chinese, which lasted about 55 minutes. The content was about the informants’ life during the Cultural Revolution in China. Their findings suggested that most of the sentences in their chosen texts, both written and oral, were in (S)VO order: 1191 out of 1270 sentences (94%) in the written text and 439 out of 477 sentences (90%) in the oral text. Furthermore, they argued that “the VO order is predominant for both definite and indefinite objects.” Therefore, Li and Thompson’s claim, namely that VO order in Mandarin Chinese signalled indefiniteness and OV order expressed definiteness, was not supported by their findings (Sun & Givon 1985, p. 344). Sun and Givon concluded that “Mandarin Chinese was a VO language and the OV order is an infrequent and marked expression” (p.344).

I would argue in support of Sun and Givon’s position that definiteness of a noun phrase in Mandarin is determined by contextual information rather than word order. Examples (59) and (60) could, in fact, have different interpretations in different contexts. For instance, the *shu* ‘book’ in example (59) would be definite if the listener had been informed that the speaker wanted to buy a particular book in a previous context. Consider (61a), a sentence uttered by speaker A in **Context I**, a made up

conversation based on example (60). Although Speaker B is informed by A in (61a) that he wanted to buy some books, B did not have any knowledge about which book Speaker A would buy. Thus, the *shu* ‘book’ which appears in (61a), (61b) and (62b) is indefinite, despite the fact that *shu* ‘book’ follows the verbs in (61a) and (61b) but precedes it in (62b).

### Context I

(61) a. Speaker A: wo xiang **mai** <sub>(verb)</sub> xie **shu** <sub>(object)</sub>.

I want buy some books.

‘I want to buy some books.’

b. Speaker B: ni xiang **mai** <sub>(verb)</sub> na ben **shu** <sub>(object)</sub>?

You want buy which class. book

‘Which book do you want to buy?’

(62) a. Speaker A: bu zhi-dao. jiu sui-bian kan-kan.

Neg. know just without purpose look

‘I don’t know. I just want to have a look.’

(Speakers A and B meet again after a period of time.)

b. Speaker A: wo **shu** <sub>(object)</sub> **mai** <sub>(verb)</sub> le.

I book buy asp.-perf.

‘I have bought a book.’

I would also agree with Sun and Givon’s argument that SOV order is a marked expression with the function of emphasizing particular information. To illustrate this

point, example (60), repeated below as (63), and (64), in which the focused information is different. Example (64), with SVO order, is merely a general statement; whereas in (63), the noun *shu* 'book' is emphasised by the use of SOV order.

(63) wo shu mai le.  
I book buy asp.-perf.  
'I have bought the book.'

(64) wo mai shu le.  
I buy book asp.-perf.  
'I have bought a book.'

The final point to be made in this section is that, in Mandarin, it is common to omit subject, object, or verb if the reference of the missing element is clear in a given context. (65) is an example in which the subject is omitted.

(65) (wo) hao xi-huan ni. (Subject is omitted.)  
(I) very like you.  
'(I) like you very much.'

Consider (66) below. In this context, speaker B omits the object *bao-gao* 'report', as it has been mentioned by speaker A.

(66) a. Speaker A: ni na zhang bao-gao da wan  
You that Mclass3 report type finish  
le ma?  
asp-perf part.-Ques.  
'Have you finished typing that report?'

b. Speaker B: wo da wan (bao-gao) le. (Object is omitted.)

I type finish (report) asp.-perf.

'I have finished typing (that report).'

In example (67), the verb *fei* 'fly' is omitted because it can be understood from the lexical meaning of *fei-ji* 'aeroplane' which often collocates with the verb *fei* 'fly', but not other verbs such as "walk", "sail", etc.

(67) Fei-ji (fei) hao kuai. (Verb is omitted)

Aeroplane (fly) very fast

'The aeroplane (is/flies) very fast.'

(68) is an example in which both subject and object are omitted. There are only two speakers in this conversation. Speaker B is the only person who can answer the question. Therefore, when B answers the question, he does not need to use "I". Similarly, the object *shu* 'book' has been mentioned by Speaker A, so it can be omitted.

(68) a. Speaker A: ni kan shu le ma?

You read book(s) asp-perf. part.-Ques.

'Have you read books?'

b. Speaker B: (wo) kan (shu) le. (Both subject and object are omitted.)

(I) read (book) asp-perf.

'(I) have read (books).'

To summarize the argument above, it is possible to find VO, SV, or only V orders in naturally occurring conversations in Mandarin. In fact, many CS data with either VO or SV word order were found in the Mandarin/Southern Min corpus. As mentioned above, this occurs because of the optional nature of arguments in Mandarin. Although the subject or object is omitted if the context provides sufficient information, it is still clear that SVO is the basic word order of Mandarin. Hence, in line with Sun and Givon's (1985) argument, it is assumed in this thesis that SVO is the basic word order of Mandarin. That assumption, with reference to the morpheme order principle of the MLF model, will be adopted in the later discussion to identify the MLs of the Mandarin/Southern Min CS data.

#### 4.2.2. Southern Min

In discussing the subject-predicate structure of Southern Min, Yang (2000) states that, in a predicate structure, the verb usually occurs before the object and after the subject and this is shown by her example in (69).

(69) wa jia bən !

I eat meal

'I eat the meal.'

(Yang 2000, p. 286)

Yang (2000, p. 286) points out that the object sometimes appears before the verb, but such a construction "serves the function of emphasizing the object". Her example is shown in (70). Yang's observation is similar to Sun and Givon's (1985) argument that



the SVO order in Mandarin is an unmarked expression while the SOV order is a marked one (c.f. section 4.2.1.).

(70) (wa) bən dʒia liao diə zao.

(I) meal eat asp-perf then leave

‘I will eat meal (first) and then leave.’

As in Mandarin, if sufficient information is provided in a given context, the subject, verb, or object can be omitted, as shown in examples (71) to (73). In (71), the subject “I” is omitted. It is clear from the context that the subject is the speaker himself, so it does not need to be mentioned.

(71) (wa) tə-ia i. (Subject is omitted)

(I) dislike him

‘(I) don’t like him.’

Consider (72), since the object *gon-kə* ‘homework’ is mentioned by Speaker A, Speaker B omits it in his response to A. Also, consider (73), since there are only two speakers in the conversation, and obviously Speaker B is the only candidate to reply to A’s question, the subject *wa* “I” can be omitted.

(72) a. Speaker A: li gon-kə jia a boe?

You homework write asp-perf. part.-Ques.

‘Have you finished writing your homework?’

b. Speaker B: wa    fia    (gon-kə)    a. (Object is omitted.)

I    write (homework) asp-perf.

‘I have written (my homework).’

(73) a. Speaker A: li    gon-kə    fia    a    boe?

You homework write asp.-perf. part.-Ques.

‘Have you finished writing your homework?’

b. Speaker B: (wa) (gon-kə)    fia    a.

(I) (homework) write asp.-perf.

‘(I) have written (my homework).’

(Both subject and object are omitted.)

Finally, the main verb in a given sentence can be omitted, where the verb is understood from the lexical meanings of the subject or object. Observe (74). The verb *zao* ‘run’ is understood from the subject *hue-chia* ‘train’ and the predicate *zo gin* ‘very fast’, so it can be omitted.

(74) hue-tʃia    (zao)    zɔ    gin. (Verb is omitted)

train    (run)    very    fast

‘The train (runs) very fast.’

From examples (69) – (74) it can be seen that the syntactic structure of Southern Min is very similar to that of Mandarin. Hence, with reference to Yang’s (2000) argument and my own observation, I argue that the basic word order in Southern Min, like

Mandarin, is also SVO.

### 4.3. Aspect

In personal communication with Myers-Scotton, she argues that the aspect markers (e.g., the perfective aspect marker “le”) are outsider late system morphemes in Mandarin. Such an argument is crucial to the present study, for the outsider late system morpheme is one of the criteria to identify the ML. The question of whether the aspect markers in Mandarin should be treated as the outsider late system morphemes will be discussed in Chapter 6. However, before trying to answer this question, it is necessary to introduce the basic functions and usages of aspect markers in Mandarin and in Southern Min.

Since most researchers into Chinese languages (Melchert 1980; Li & Thompson 1981; Norman 1988; Hu, Pan & Xu 2001; Yip & Don 2004) agree that Chinese does not have tense, but does have aspect, this section will then focus on “aspect” in Mandarin and Southern Min. According to Li and Thompson (1981), there are four kinds of aspect in Mandarin, namely perfective, durative, experiential, and delimitative. In the literature on Southern Min, the issue aspect has rarely been discussed. However, I would adopt Li and Thompson’s (1981) perspective and argue that Southern Min also has four kinds of aspects since the syntax of Southern Min is identical to Mandarin.

#### 4.3.1. The Perfective Aspect

##### 4.3.1.1. Mandarin Perfective Aspect Marker: “le”

Li and Thompson (1981) point out that the perfective aspect in Mandarin is expressed through the particle “le”, which occurs after the verb. To illustrate the function of “le” and show how its use changes the meaning of the sentences, I provide examples in

(75). (75a) is simply a statement of an event. With the addition of the perfective aspect marker “le”, (75b) expresses the completion of an event. Very often, the use of “le” denotes that the reported event has been mentioned earlier in the context. This means that a listener to an utterance like (75b) may have previous knowledge of Mei-lin’s writing a report.

(75) a. Mei-lin xie bao-gao.

Name write report.

‘Mei-lin writes a report.’

b. Mei-lin xie **le** bao-gao.

Name write asp.-perf. report

‘Mei-lin finishes writing the report.’

It is important to note that a distinction has to be made between the aspect marker “le” and the sentence final particle “le”. The aspect marker “le” as stated above signals grammatical information while the sentence final particle ‘le’ has specific discourse functions. If the main verb of a given sentence is intransitive as shown in (76a), the aspect marker “le” will also occur in the sentence final position. (76b) is Yip and Don’s (2004, p.319) example showing one of the functions that the sentence final particle “le” serves, that is “a speaker’s response to a situation that is markedly better or worse than expected”. We can only differentiate these two usages from the context.

(76) a. ka-che bao-zha **le**.

Truck explode asp.-perf.

‘The truck has exploded.’

b. wo-men tai xing-yun **le**.

We        very    lucky        part.-unexpected

‘We are really lucky!’

#### 4.3.1.2. Southern Min Perfective Aspect Markers: “liao” and “a”

According to Kho (1998), in Southern Min the particle “liao” is added after the verb in order to express a completion of an action. I argue that the function of “liao” is exactly the same as the Mandarin perfective aspect marker “le”. Consider my examples (77) to see how “liao” is used.

(77) a. in        tsa        am-dən.

They    cook    dinner

‘They cook dinner.’

b. in        tsa        **liao**        am-dən.

They    cook    asp.-perf. dinner

‘They finish cooking dinner.’

In addition to “liao”, Chen (1989) argues that in Southern Min the perfective aspect could also be marked by the sentence final particle “a”. I agree that “a” has the same function as “liao”, but they occur at different positions in sentences, “liao” occurs after the main verb and “a” occurs in the sentence final position. Consider my examples in (78).

(78) a. wa    ʃia        gon-kə.

I        write    homework

‘I do the homework.’

b. wa jia **liao** gon-kə.

I write asp-perf homework

‘I finish doing the homework.’

c. wa jia gon-kə **a**.

I write homework asp-perf

‘I finish doing the homework.’

#### 4.3.2. The Durative Aspect

##### 4.3.2.1. Mandarin Durative Aspect Markers: “zai” and “zhe”

According to Crystal (1997, p.128), the durative aspect refers to “an event involving a period of time”, and is often known as “continuous” or “progressive” in grammatical analysis. There are two durative aspect markers in Mandarin, namely “zai” and “zhe”. These two particles carry the same grammatical information, but their position in the syntactic structure is rather different. “zai” only occurs in front of the verb. In order to illustrate the function of “zai”, I provide my own example with no aspect markers (79a) in order to make a comparison with (79b), an example cited from Li and Thompson (1981, p.218).

(79) a. Lisi jie-shi wen-fa.

Lisi explain grammar

‘Lisi explains the grammar.’

b. Lisi **zai** jie-shi wen-fa.

Lisi asp-dur explain grammar

‘Lisi is explaining the grammar.’

The particle “zhe”, however, appears only after the verb, and this is shown in my own example (80).

- (80) Lisi jie-shi **zhe** wen-fa.  
Lisi explain asp-dur grammar  
‘Lisi is explaining the grammar.’

In addition, “zai” and “zhe” can also appear in the same sentence, and this is shown in my example (81). The co-occurrence of these two has no effect on meaning.

- (81) Lisi **zai** jie-shi **zhe** wen-fa.  
Lisi asp-dur explain asp-dur grammar  
‘Lisi is explaining the grammar.’

#### 4.3.2.2. Southern Min Durative Aspect Marker: “di”

According to Kho (1998), the particle “di” (also known as “le” or “de” in some varieties of Southern Min) expresses an action which is in progress, and occurs before the verb. (82) is my example showing how “di” is used.

- (82) a. sen-sei gai-sue bun-hua  
teacher explain grammar  
‘The teacher explains/explained the grammar.’  
b. sen-sei **di** gai-sue bun-hua.  
teacher asp-durI explain grammar  
‘The teacher is explaining the grammar.’

### 4.3.3. The Experiential Aspect

#### 4.3.3.1. Mandarin Experiential Aspect Marker: “guo”

Li and Thompson (1981, p.226) suggest that the aspect marker “guo” refers to “an event that has been experienced at least once at some indefinite time”. As a contrast with Yip and Don’s (2004, p.109) example of “guo” in (83b), I also list (83a), a sentence with no aspect marker.

(83) a. wo kan na ben xiao-shuo.

I read that Mclass2 novel

‘I read that novel.’

b. wo kan **guo** na ben xiao-shuo.

I read asp-exp that Mclass2 novel

‘I have read that novel.’

#### 4.3.3.2. Southern Min Experiential Aspect Marker: “gue”

Yang (2000) points out that the experiential particle in Southern Min, which is equivalent to “guo” in Mandarin, is “gue”. Compare my examples in (84) to see how “gue” is used.

(84) a. i ki Dai-ba.

he go Taipei

‘He goes/went to Taipei.’

b. i ki **gue** Dai-ba.

he go asp-exp Taipei

‘He has been to Taipei.’



#### 4.3.4 The Delimitative Aspect

##### 4.3.4.1. Mandarin Delimitative Aspect

The last kind of aspect discussed by Li and Thompson (1981, p.232) is the delimitative aspect which refers to “doing an action a little bit or for a short period of time”. This kind of aspect is expressed through the reduplication of the verb. Their examples are shown in (85).

(85) a. ni **shi shi** kan.

you try try see

‘Try it just a little and see.’

(Li & Thompson 1981, p.232)

b. ta **shui shui** jiu hao.

S/he sleep sleep then well

‘S/he will be well after just sleeping a little.’

(Li & Thompson 1981, p.233)

##### 4.3.4.2. Southern Min Delimitative Aspect Marker: “lei”

Chen (1989) argues that one way to mark the delimitative aspect in Southern Min is to add the particle “lei”. To illustrate this, observe her example in (86).

(86) dan ji **lei!**

wait one asp.-delim

‘Wait just for a moment!’

(Chen 1989, p.69)

It should be noted that there is no Mandarin particle with the same function as the

“lei” in Southern Min. However, as Chen points out, sometimes the delimitative aspect in Southern Min, similar to Mandarin, can also be marked by the construction: “reduplication of the verb + lei”. Observe the example in (87).

- (87) yi    **kuen**   **kuen**   **lei**            dio    ho.  
      he   sleep   sleep   asp-delim   then   fine  
      ‘He will be fine after just sleeping a little.’

#### 4.4. Classifiers

An important feature of Chinese is the use of classifiers, which are used to modify nouns. Li and Thompson (1981, p.105) point out that “the choice of classifier is determined by the noun”. Hence, in choosing an appropriate classifier, one considers the shape, size, animacy or other relevant features of the noun being modified. Their definition also holds for the usage of the Southern Min classifiers. The point of discussing classifiers in this section and in section 4.5 the possessive markers and the nominalizer is that I argue that they constitute early system morphemes and bridge late system morphemes with reference to Myers-Scotton’s (2002a) 4-M model. The problematic application of the morpheme order principle and the system morpheme principle of the MLF model to the Mandarin/Southern Min CS data will be discussed in Chapter 6. Other possible criteria to identify the ML will be required. Examining the early system morphemes and bridge late system morphemes in the identification of the ML will be one of the additional criteria proposed by the present study. Hence, it is crucial to introduce the functions of classifiers, possessive markers, and the nominalizer in this chapter.

In this section, I will focus on demonstrating the similarities between the usage of the

classifiers in Mandarin and Southern Min. Table 1 is a list of glosses of Mandarin and Southern Min classifiers, which are used in this thesis. It should be noted that Table 1 only includes a very small number of classifiers in Mandarin and Southern Min. The classifier systems in these two languages are much more complex and are beyond the scope of this study.

**Table 1: Mandarin and Southern Min Classifiers with Glosses in English**

Semantic Features	Mandarin	Gloss	Southern Min	Gloss
long and round	zhi	Mclass 1	gi	Sclass 1
square and thick	ben	Mclass 2	bun	Sclass 2
square and thin	zhang	Mclass 3	diun	Sclass 3
round body	ke	Mclass 4	liap	Sclass 4
human	ming	Mclass5	mia	Sclass5
animal	tiao	Mclass6	dʒia	Sclass6
building	jian	Mclass7	gin	Sclass7
vehicles on the land	tai	Mclass8	dar	Sclass8
long construction	dao	Mclass9	də	Sclass9
event	hui	Mclass10	giam	Sclass10

To see the function of these classifiers, let us observe the examples below that are cited in Yip and Don (2004, p.33). In the Mandarin example in (88a), the classifier “zhi” is used since the shape of the noun *bi* ‘pen’ is long and round. For the same reason, the classifier “gi” is used in the Southern Min example in (88b).

(88) a. *yi zhi bi.* (Mandarin)

one Mclass1 pen

‘one pen’

b. *dʒɪ gi bɪ* (Southern Min)

one Sclass1 pen

‘one pen’

(*Italic =Mandarin*; **Bold = Southern Min**)

Also consider the examples in (89) and (90). In the Mandarin example in (89a) and (90a), the shape of either *zhi* ‘paper’ or *shu* ‘book’ is rectangular, so the classifiers chosen should reflect this feature. Furthermore, the thickness of the two objects also needs to be considered. The classifier *zhang* is used to describe something rectangular and thin. Hence, it is selected to describe *zhi* ‘paper’ in (89a). The classifier *ben* is used to describe something rectangular and thick. Therefore, it is used to modify *shu* ‘book’ in (90a). This same analysis holds for the Southern Min examples in (90a) and (90b).

(89) a. *yi zhang zhi* (Mandarin)

one Mclass3. paper

‘one piece of paper’

b. **dzɿ diun zwa.** (Southern Min)

one Sclass3 paper

‘one piece of paper.’

(90) a. *yi ben shu.* (Mandarin)

one Mclass2 book

‘one book’

b. **dzɿ bun tse.** (Southern Min)

one Sclass2 book

‘one book.’

(*Italic* = Mandarin; **Bold** = Southern Min)

Further examples of classifiers are provided in (91) to (93).

(91) a. *yi* *ming* *jun-ren*. (classifiers which modify human beings)

one Mclass5 soldier

‘one soldier’

b. **ɔ̃ɰ** **mia** **gun-lin**

one Sclass5 soldier

(92) a. *yi* *tiao* *go* (classifiers which modify animals)

one Mclass6 dog

‘one dog’

b. **ɔ̃ɰ** **ɔ̃ɰia** **gaʊ**

one Sclass6 dog

‘one dog’

(93) a. *yi* *jian* *fang-zi* (classifiers which modify building)

one Mclass7 house

‘one house’

b. **ɔ̃ɰ** **gin** **tsu**

one Sclass7 house

‘one house’

(*Italic* = Mandarin; **Bold** = Southern Min)

Another point that needs to be raised here is that “a classifier must occur with a number, demonstrative or certain quantifiers” (Li & Thompson 1981, p.104). This is referred to by Li and Thompson (1981, p.104) as a “classifier phrase”, whose construction is shown in (94).

(94) Numeral/demonstrative/quantifier + classifier + noun

A classifier must co-occur with either a numeral or a demonstrative, or a quantifier, otherwise it is not grammatical. (88) to (93) are examples of numeral + classifier + noun. (95) to (97) include some examples of my own which illustrate the use of demonstratives and quantifiers in the construction shown in (94). The two examples in (98) are grammatically incorrect because no numerals, demonstratives, or quantifiers occur before the classifiers.

(95) a. *na ming jun-ren.*

that Mclass5 soldier

‘That soldier’

b. **hi mia gun-lin**

that Sclass5 soldier

‘That soldier’

(96) a. *ji ming jun-ren.*

Several Mclass5 soldier

‘Several soldiers’

b. **gui**      **mia**      **gun-lin**

several    Sclass5    soldier

‘several soldiers’

(97) a. *na* *ji*      *ming*      *jun-ren.*

that    several    Mclass5    soldier

‘Those soldiers’

b. **hi** **gui**      **mia**      **gun-lin**

that    several    Sclass5    soldier

‘Those soldiers’

\*(98) a. *ming*      *jun-ren.*

Mclass5    soldier

b. **mia**      **gun-lin**

Sclass5    soldier

(*Italic = Mandarin; Bold = Southern Min*)

A classifier, which is not followed by a noun, can refer to the noun occurring in the classifier phrase of a preceding utterance. This is illustrated by my own Mandarin and Southern Min examples in (99) and (100).

(99) a. Speaker A: *wo* *yao* *mai* *yi tai* *che.* (Mandarin)

I    will    buy    one    Mclass8    car

‘I will buy a car.’

b. Speaker B: *wo* *yie* *yao* *mai* *yi tai.*

I    also    will    buy    one    Mclass8

‘I will buy one, too.’

(100)a. Speaker A: **wa bē bè dʒ<sub>1</sub> da<sub>1</sub> tʃia.** (Southern Min)

I will buy one Sclass8 car

‘I will buy a car.’

b. Speaker B: **wa ma bē bè dʒ<sub>1</sub> da<sub>1</sub>.**

I also will buy one Sclass8

‘I will buy one, too.’

#### 4.5. Possessive Marker and Nominalizer

##### 4.5.1. Mandarin

The particle “de” in Mandarin has two different functions, namely to mark possessiveness and to nominalise a verb phrase, turning it into a relative clause. In the following sections, we will discuss these two functions in more detail.

##### 4.5.1.1. Possessive Marker

In Mandarin, the concept of possession is marked by using the genitive/possessive particle “de” whose function is similar to the possessive marker “ ’s ” or “of” in English. For instance:

(101) lao-shi **de** nu-er

teacher poss. daughter

‘teacher’s daughter / daughter of teacher’



#### 4.5.1.2. Nominalizer

Li and Thompson (1981) define nominalization as a grammatical process that turns a verb, or a verb phrase into a noun phrase. They argue that in Mandarin that process is accomplished by adding the particle “de” after the verb. To illustrate this point, see my examples in (102) and (103).

(102)a. *he* (verb)

drink

b. *he de* (noun)

drinks (e.g., coke, tea etc.)

(103)a. *zhong shui-guo* VP

grow fruit

b. *zhong shui-guo de*. NP

grow fruit nom

‘a/the person who grows fruit.’

In (102), with the addition of the nominalizer “de”, the verb *he* ‘drink’ becomes a noun *he de* ‘drink’. The verb phrase *zhong shui-guo* ‘grow fruit’ in (103a) turns into a noun phrase in (103b) after the nominalizer “de” has been added. Li and Thompson also point out that the addition of the nominalizer “de” can turn a sentence into a subordinate clause which functions as a modifier. Consider my examples in (104).

(104)a. *wo mai za-zhi*.

I buy magazine

‘I bought a magazine.’

b. [wo mai za-zhi **de**] shu-dian zai lu di.



I buy magazine nom bookstore just road bottom

‘The bookstore where I bought the magazine is at the end of this road.’

According to Li and Thompson’s argument, the nominalizer “de” turns the sentence in (104a) into a subordinate clause in (104b) which modifies the noun *shu-dian* ‘bookstore’. The English translation of the elements in the bracket in (104b) may be a subordinate clause, but in Mandarin they act as an NP modifier. Hence, I would argue that a nominalizer in Mandarin also has the function of connecting a sentence and a noun phrase.

#### 4.5.2. Southern Min

##### 4.5.2.1. Possessive Marker

I have shown that the particle “de” in Mandarin serves the function of marking possession. The equivalent particle in Southern Min is “e”. To illustrate its usage see Yang’s (2000, p.282) examples in (105).

(105)a. a-gon e ba-giã.

grandfather poss glasses

‘Grandfather’s glasses’

b. to-a e ba-dʒiu.

rabbit poss. eye

‘The rabbit’s eyes’

#### 4.5.2.2. Nominalizer

I argue that the nominalizer “e” in Southern Min has the same function as the “de” in Mandarin, namely turning a verb or a verb phrase into a noun phrase. Consider my examples in (106) and (107).

(106)a. lim (verb)

drink

b. lim e (noun)

drinks (e.g., coke, tea, etc.)

(107)a. bei tsai <sub>VP</sub>

sell vegetables

b. bei tsai \_\_\_\_\_ e <sub>NP</sub>

sell vegetables nom

‘the person who sells vegetables’

Moreover, the Southern Min nominalizer “e” can also turn a sentence into an NP modifier to modify a head noun. Consider the examples in (108).

(108)a. wa jie tse.

I borrow book

‘I borrowed a book.’

b. [wa dʒiə tse e] so-zai dio ʃi dʒia.



I borrow book nom place then cop. here

‘This is the place where I borrowed the book.’

The English translation of the elements in the bracket in (108b) may be a subordinate clause, but in Southern Min they act as an NP modifier. Hence, I also argue a nominalizer in Southern Min also has the function of connecting a sentence and a noun phrase.

The main reason to introduce the functions of the possessive markers and nominalizers in Mandarin and Southern Min is that they are one of a type of morphemes specified in Myers-Scotton (2002a)’s MLF model, namely bridge morphemes (c.f. section 3.3.2). Although bridge morphemes are not crucial in the identification of the ML in the original version of MLF model, they, however, will become important in the discussion in section 6.2 in which I seek new criteria to solve the theoretical problems of this model. .

#### 4.6. Serial Verbs

One of the CS models this thesis aims to test, namely the MLF model, has a bilingual clause as its basic unit of analysis. Hence, the issue of how to identify a clause is crucial to this thesis. In English, we can identify a clause based on the main verb, for only one main verb is allowed in a clause. Both Mandarin and Southern Min have the construction of serial verbs<sup>14</sup>, and therefore identifying a clause is more complex. This section will provide a brief description of serial verbs in Mandarin and Southern Min.

According to Li and Thompson (1981, p.594), the serial verb construction in

---

<sup>14</sup> According to Crystal (1997, p.348), “Serial verb refers to a type of construction for a sequence of verbs or verb phrases within a clause (or a sequence of clauses)... These verbs share a semantic argument, but there is no conjunction or inflection to mark co-ordination or subordination”.

Mandarin refers to “a sentence that contains two or more verb phrase or clauses juxtaposed without any marker indicating what the relationship is between them.” To my knowledge, there is almost no discussion of the issue of “serial verbs” in the literature on Southern Min. Since Southern Min shares most of its syntax with Mandarin, I argue that Li and Thompson’s (1981) proposals with respect to the usage of the serial verbs in Mandarin hold for the analysis of those in Southern Min. Consider their Mandarin example in (109) and my Southern Min examples in (110).

(109) *ta tian-tian chang ge xie xin.* (Mandarin)

3s day-day sing song write letter

‘Everyday, s/he sings songs and writes letters.’

(Li and Thompson 1981, p.595)

(110) *i mui-gan tjiu gua fia pue.* (Southern Min)

3s everyday sing song write letter

‘Everyday, she sings songs and writes letters.’

At first sight, the Mandarin sentence in (109) has two main verbs. Li and Thompson suggest that the relationship between these two verb phrases, namely *chang ge* ‘sing songs’ and *xie xin* ‘write letters’, is consecutive. In other words, one occurs right after the other, though there is no overt marker (i.e., the coordinating conjunction) to mark such a relationship. The same analysis holds for the Southern Min example in (110) since it has exactly the same syntactic structure as the Mandarin example in (109). I argue that a sentence, which has a serial verb construction like (109) or (110), would be considered to have two clauses. This is because the two verb phrases indicate two

separate events and have a coordinative relationship.

In addition to Li and Thompson's (1981) argument, Li (1990) points out that there is another kind of serial verb construction in which the two verbs share the same meaning. To illustrate this point, my own Mandarin examples are provided in (111).

(111) a. *shi-bing* *jian-zhu* *le* *yi* *dao* *ho* *qiang*.  
soldier build-construct asp-perf one Mclass9 thick wall.

'The soldiers built a thick wall.'

b. *shi-bing* *jian* *le* *yi* *dao* *ho* *qiang*.  
soldier build asp-perf one Mclass9 thick wall.

'The soldiers built a thick wall.'

c. *shi-bing* *zhu* *le* *yi* *dao* *ho* *qiang*.  
soldier construct asp-perf one Mclass9 thick wall.

'The soldiers built a thick wall.'

The underlined verbs *jian* 'build/construct' and *zhu* 'build/construct' in (111a) have almost the same meaning. If one of them is omitted, as (111b) & (111c) show, the meaning of the entire sentence does not change. In parallel to Li's (1990) observation in Mandarin, I argue that Southern Min also has similar serial verb construction as shown by (111a). Hence, the analysis of the Mandarin examples in (111) holds for the Southern Min examples in (112).

(112) a. *lan* *fj-i-am* *kuan-mai*.

we test see

‘Let’s test it and see.’

b. **lan** **fj** **kuan-mai**.

we test see

‘Let’s test it and see.’

c. **lan** **iam** **kuan-mai**.

we test see

‘Let’s test it and see.’

In the present study, a sentence with serial verb construction like (111a) or (112a) would be considered to have only one clause. This is because the meanings of the verbs are identical, and the absence of one of the verbs does not change the meaning of the entire clause. This criterion will be adopted to identify what a bilingual Mandarin/Southern Min clause is, if a given bilingual utterance in the CS corpus has a similar serial verb construction to those illustrated by the examples in (109) – (112).

#### 4.7. Sentence Final Particles

In their research into English/Chinese CS in Singapore, Kamwangamalu and Li (1991) adopted a lexical and structural approach, particularly by using the criteria of word order and some syntactic elements to explore the issue of matrix language assignment. Since there is a lack of inflectional morphology in Chinese, they proposed some new criteria (e.g., sentence final particles and verbless clauses) to identify the matrix language in their CS data. In section 6.2.2, the present study will show that the two criteria, namely the morpheme order principle and the system morpheme principle, in the MLF model are found not to be applicable to the Mandarin/Southern Min CS data. In this case, the criteria (e.g., checking the sentence final particles) proposed by

Kamwangamalu and Li (1991) to identify the MLs are likely to solve the problem encountered in the present study. Hence, it is relevant to have a special section discussing the issue of sentence final particles in Mandarin and Southern Min.

There are many sentence final particles in Mandarin and Southern Min, and each of them serves different semantic and pragmatic functions. However, I will only introduce the sentence final particles and their functions that occur most frequently in the CS data collected from my field research.

#### 4.7.1. The Particle “a”

##### 4.7.1.1. Mandarin

The sentence final particle “a” has several different functions, namely a marker of exclamations, exclamatory questions, reducing forcefulness and warning. Each function will be described in more detail below.

##### *Exclamation Marker*

The particle “a” is often to signify an exclamation. Consider my examples in (113) and (114).

- (113) wu-zi li duo an-jing a!  
house in much quiet part.-excl.  
‘How quiet it is in the room.’

- (114) duo ban a!  
much wonderful part.-excl.  
‘How wonderful!’



### *Exclamatory Question Marker*

Yip and Don (2004, p.357) argue that the particle “a” often occurs in an exclamatory question that “expresses surprise, doubt etc.” (115) is the example they give to illustrate the function of expressing surprise.

- (115) zhe shi zen-me hui shi a?  
this cop. what Mclass10 affair part.- EQM  
‘What on earth is going on?’  
(Yip & Don 2004, p.357)

### *Reducing Forcefulness*

The particle “a” is often used to reduce the forcefulness or to soften the tone of an utterance. Consider Li and Thompson’s (1981, p.315) example in (116). (116a) is a command or order, but with the addition of the particle “a”, the forcefulness is reduced and becomes more like a suggestion as shown in (116b).

- (116)a. ni lai!  
you come  
‘You come here!’  
b. ni lai a.  
you come part.-RF  
‘You come here.’  
(Li & Thompson 1981, p.315)

### *Warning*

One more function that the particle “a” has is that it signals the speaker’s warning to the addressee, as shown in (117).

- (117) zhei ren de hua shi kao bu zhu de a.  
this person poss words cop rely neg. hold nom. part.-warn  
‘This person’s words are unreliable.’  
(Li & Thompson 1981, p. 316)

### *Agreement Marker*

A last function that the Mandarin sentence final particle “a” has is to mark the listener’s agreement with the speaker’s statement or comments. Consider the examples in (118).

- (118)a. Speaker A: ri-ben chi de pin-zhi hao.  
Japan car poss quality good  
‘The quality of Japanese cars is good.’  
b. Speaker B: due a! due a!  
Yes part.-agree Yes part.-agree

#### 4.7.1.2. Southern Min

In section 4.3.1.2, I stated that the Southern Min sentence final particle *a* acts as a perfective aspect marker. This section will introduce other functions of this particle.

### *Agreement Marker*

Li (1999) argues that the particle *a* serves the function of signalling the listener’s

agreement with the speaker's opinion. This function overlaps that of the Mandarin sentence particle "a". Consider my examples in (119) and (120).

(119)a. Speaker A: li-bun e      ʈʂia   ʃi-zai   ping-dʒi   ho.  
Japan poss car really quality good  
'The quality of Japanese cars is really good.'

b. Speaker B: ʃi   a!                  ʃi   a!  
yes part.-agree. Yes part.-agree  
'Yes! Yes!'

(120)a. Speaker A: i   kə-ʈʂi   zo   ʈʂi-tsam!  
he test very bad  
'The result of his test was very bad.'

b. Speaker B: dio   a!  
right part.-agree.  
'That's right!'

### *Explanation*

The other function of the particle *a* suggested by Li (1999) is to provide an explanation in answer to a "why" question. Observe my examples in (121) and (122).

(121)a. Speaker A: wi-ʃia-mi li   e-hiao   gon   in-gi?  
why you can speak English  
'Why can you speak English?'

b. Speaker B: wa di in-go ta dai-ha e a.

I in UK study university nom. part.-expl.

‘I studied at university in the UK.’

(122)a. Speaker A: wi-fia-mi dʒia-ni gan-dan e de-bo ma

Why such easy nom question also

be-hiao zo.

neg-can answer

‘Why can’t you answer such an easy question?’

b. Speaker B: i bo ga wa a.

he neg. teach me part.-expl.

‘He didn’t teach me.’

### *Warning*

I argue that the particle *a* in Southern Min also has the function of expressing a speaker’s warning to an addressee. This function is again identical to that of the particle *a* in Mandarin.

(123)li diə-ai ʃə-fim a! (Southern Min)

you should careful part.-warn.

‘You should be careful!’

### *Exclamatory Question Marker*

The particle “a” in Mandarin and the particle *a* in Southern Min share one more function, namely as an exclamatory question marker. Consider my example in (124).

- (124) dʒin-ma ʃi      hua-ʃin      ʃia-mi      dai- dʒi a?  
 now      cop      happen      what      affair      part.-EQM  
 ‘What on earth is going on?’

#### 4.7.2. The Particle “la”

##### 4.7.2.1. Mandarin

###### *Occurrence of Unfortunate Events*

Yip and Don (2004) suggest that “la” is used to express the fact that something important and unfortunate has happened. Consider my examples in (125).

- (125)a. ta      sha ren la!  
 3s      kill person      part.- unfor  
 ‘S/he just killed a person!’
- b. di-ren      lai      la!  
 enemy      come      part.-unfor  
 ‘The enemies have come!’

###### *Marking a Speaker’s Impatience*

The sentence final particle “la” has the other function, namely to mark a speaker’s impatience. Consider my examples in (126).

- (126)a. wo bu      zhi-dao la!  
 I      neg know      part.-impat  
 ‘I don’t know!’ (And I resent you keep asking!)

- b. bei lao wen wo la!  
 neg again ask me part.-impat  
 ‘Don’t ask me all the time!’

#### 4.7.2.2. Southern Min

##### *Occurrence of Unfortunate Events*

According to Chen (1989), the particle *la* in Southern Min has the same function as the “la” in Mandarin, namely to mark the occurrence of some unfortunate events. To illustrate this, I provide my examples in (127).

- (127)a. i si la!  
 She/he die part.-unfor  
 ‘She is dead!’
- b. de-lin lai a la!  
 enemy come asp-perf part.-unfor  
 ‘The enemies have come!’

##### *Marking a Speaker’s Impatience*

Chen (1989) goes on to suggest that the Southern Min sentence final particle *la* can also be used to mark the speaker’s impatience with the addressee. Again, this is identical to the function of the “la” in Mandarin. Consider her example in (128), and my example in (129).

- (128) wa mu zai la!  
 I neg. know part.-impat.

'I don't know! (And I resent that you keep asking!)

(Chen 1989, p.53)

(129) mai gə mən la!

Neg. again ask part.-impat.

'Don't ask my again! (I feel annoyed that you keep asking!)

#### 4.7.3. The Question Particle "ma"

##### 4.7.3.1. Mandarin

The particle "ma" in Mandarin is a question marker that can turn a declarative sentence into a question (Li & Thompson 1981). This is illustrated by my example in (130). With the addition of the particle "ma", the statement (130a) is turned into a Yes-No question.

(130)a. ta ai Yun-Fan.

3s love Name

'S/he loves Yun-Fan.'

b. ta ai Yun-Fan ma?

3s love Name part.-Ques

'Does s/he love Yun-Fan?'

##### 4.7.3.2. Southern Min

Similar to the particle "ma" in Mandarin, the particle *ma* in Southern Min can turn a statement into a yes-no question. Observe my examples in (131).

(131)a. A-min ki chia-zam.

Name go station

'A-min goes to the station.'

b. A-min ki chia-zam ma?

Name go station part.-Qques

'Does A-min go to the station?'

In summary, the different functions of some Mandarin and Southern Min sentence final particles, which occurred most frequently in my CS data, were reviewed in this section. The reason for this discussion is that Kamwangamalu and Li (1991) used sentence final particles in Chinese as one of the criteria to identify the ML of their English/Chinese CS data. Since this criterion will be tested with the Mandarin/Southern Min data (c.f. section 6.2.2.3), it is, then, important to describe the similarities and differences between the usages of Mandarin and Southern Min sentence final particles.

#### 4.8. The *A-not-A* Question vs. *A ...not* Question Construction

##### 4.8.1. The *A-not-A* Question in Mandarin

An A-not-A question is a kind of disjunctive question whose function is "to present an either-or choice to the respondent" (Li & Thompson 1981, p.531). The general construction for this type of question is to combine both the affirmative and the negative version in one sentence. The construction of the A-not-A question is: verb + negation + verb. To illustrate this, see my examples in (132).

(132)a. ta yao qu kan dian-yin. (affirmative)

S/he want go see movie



‘S/he wants to see a movie.’

- b. ta bu yao qu kan dian-yin. (negative)

S/he neg. want go see movie

‘S/he does not want to see a movie.’

- c. ta yao bu yao qu kan dian-yin?

S/he want neg. want go see movie

‘Does s/he want to see a movie or not?’

(132a) is a sentence which expresses affirmation. As shown in (132b) the addition of the negation marker *bu* ‘not’ turns (132a) into a negative sentence. (132c) is an example with the A-not-A construction, which includes both the affirmative and negative expression. The question in (132c) proposes an either-or choice, namely “she wants to see a movie” or “she doesn’t want to see a movie”, for the listener to answer.

#### 4.8.2. The *A ...not* Question in Southern Min

In the previous section, the A-not-A question in Mandarin was discussed. I argue that Southern Min has an equivalent expression. This is illustrated by my examples in (133).

- (133)a. Ahon be ki bɛn-in. (affirmative)

Name want go hospital

‘Ahon wants to go to the hospital.’

- b. Ahon bo be ki bɛn-in. (negative)

Name neg. want go hospital

‘Ahon does not want to go to the hospital.’

c. Ahon be ki bɛn-in bo?

Name want go hospital neg.

‘Does Ahon want to go to the hospital or not?’

Compare (132) and (133). We can see that although a Mandarin A-not-A question has the same function as a Southern Min A...Not question, there is a difference in the syntactic structure. In the Mandarin example (132c), the negative version *bu yao* ‘neg. + want’ occurs immediately after its affirmative version *yao* ‘want’. However, in the Southern Min example in (133c) the negative version occurs in the sentence final position. Hence, the construction of the A-not-A question in Southern Min is: “verb + ... + negation”. The reason for making such a comparison is that at the beginning of this chapter I mentioned that the major differences between Mandarin and Southern Min are at phonological and lexical levels. The A-not-A questions in Mandarin and the A...Not questions in Southern Min are the very few examples in my data, which show different syntactic constructions. According to the MLF model, surface word order is an important criterion to identify the ML. Thus, the discussion in this section will be helpful to identify the ML of a bilingual Mandarin/Southern Min clause.

#### 4.9. A Brief Description of Tsou Grammar

The study of Tsou grammar is still at a very early stage. Hence, the relevant literature (for example, Zeitoun 1992; Chang 1998; Chang 2004) is very limited. The grammatical functions of many linguistic elements in Tsou, such as the case marking system, the tense and aspect system, and the thematic role marking system have been discussed in the literature, but are still open for discussion. The linguists who study Tsou seem to have different views regarding its grammatical system. For instance,

Zeitoun (1992) argues that the tense of Tsou is divided into past and non-past, while Chang (1998) argues that Tsou tense should be divided into future and non-future. Moreover, it is frequently found that the arguments proposed by individual scholars are not consistent. For example, Chang (1998, p.42) first claimed that “Tsou tense can be broken down into [+future] and [-future]”, but later (p.43) she claimed that “Tsou tense system consists of three tenses – the present, the past and the future.” In her later research, Chang (2004, p.10) completely changes her perspective and claims that “Tsou tense seems not to be expressed in morphosyntactic categories... Tsou is a language without structured tense”. Although these issues present difficulties in the research into Tsou, I will try to draw a general picture of the grammatical system of Tsou.

As has been mentioned repeatedly, word order and inflectional morphology are two important criteria in identifying the ML in Myers-Scotton’s (2002a) MLF model. Thus, this section will focus on an introduction to the basic word order and inflectional morphology of Tsou. The organisation of this section is as follows: First, the case marking system of Tsou will be introduced. Second, a discussion of its basic word order will be provided. Third, I will briefly discuss the tense and thematic role marking system.

#### 4.9.1. The Case Marking System

According to Zeitoun (1992) and Chang (1998), case markers in Tsou are divided into two major categories, namely nominative case markers and oblique case markers. The nominative case marker marks the subject of a given sentence, while the oblique case markers mark the non-subject noun phrases. Consider the example in (134), which is a monolingual Tsou sentence collected from my fieldwork. In (134), the subject

*tsəmoi* ‘bear’ is marked by the nominative case marker *?o*. The object *oko* ‘child’ is marked by the oblique case marker *da*.

- (134) mo            ahdu   bohtsinhi   da   oko   ?o   tsəmoi.  
Tens3,Agent   neg.   attack   Ob1. child   Nom4 bear  
‘The bear did not attack the child.’

The complete case marking system in Tsou is very complex, and each case marker carries an abundance of semantic information. In order to provide a general picture about how the system works and the semantic information these case markers carry, Table 2 is given below. It is a modified version of Chang’s (1998, p.20) table with the addition of the “Gloss” column in order to label each case marker.

**Table 2: The Semantic Features of Tsou Case Markers**

	Case Marker	Gloss	[vis]	[far]	[near]	[i/f]	[seen]	[heard /smelt]	[loc]
<b>Nominative</b>	<b>?e</b>	<b>Nom1</b>	+	-	+		+		+
	<b>si</b>	<b>Nom2</b>	+		-	+	+		+
	<b>da</b>	<b>Nom3</b>	+	+	-		+		+
	<b>?o</b>	<b>Nom4</b>	-	+	-		+		+
	<b>na</b>	<b>Nom5</b>	-			-	-		+
	<b>co</b>	<b>Nom6</b>						+	
	<b>ne</b>	<b>Nom7</b>							+
<b>Oblique</b>	<b>da</b>	<b>Obl1</b>	+	-		+	+		+
	<b>do</b>	<b>Obl2</b>	-	+			+		+
	<b>no</b>	<b>Obl3</b>	-			-	-		+
	<b>co</b>	<b>Obl4</b>						+	

[vis]= visible to the speaker (at the time of making the utterance)

[far]=far from the speaker

[near]=near the speaker

[i/f]=intimate/familiar to the speaker

[seen]=seen by the speaker before

[heard/smelt]=audible or can be smelt by the speaker

[loc]=locative

‘+’ = the presence of a feature; ‘-’= absence of a feature;

blank = the feature is irrelevant

In the MLF model, case affixation is an important criterion (i.e., the system morpheme principle) (c.f. section 3.3), which can be used to identify the ML. Thus, it is necessary to introduce the functions of the case affixes in Tsou. As shown by Table 2, the complete Tsou case marking system is very complex. The usage of only several case markers here will be explained to provide a general idea.

4.9.1.1. The Feature “[+/- vis]”: “?e” vs. “?o” and “da” vs. “do”

According to Chang’s (1998) table, the major difference between the two nominative case markers, namely *?e* and *?o*, is that *?e* indicates that the subject was visible to the speaker at the time of the utterance was made, while *?o* indicates that the subject was not visible. Similarly, the two oblique case markers, namely *da* and *do*, mark the visibility and non-visibility of the object *oko* ‘child’. Compare Zeitoun’s (1992, p.75) examples in (135a) and (135b).

(135)a. moso            eobako da            oko    ?e            ino  
 tense1,Agent beat    Obl1,**[+vis]**    child    Nom1,**[+vis]**    mother

‘The mother hit the child.’ (Both the mother and the child were visible to the speaker of this utterance.)

b. moso            eobako do            oko    ?o            ino  
 tense1,Agent beat    Obl2,**[-vis]**    child    Nom4,**[-vis]**    mother

‘The mother hit the child.’ (Neither the mother nor the child was visible to the speaker of this utterance.)

In (135a), both the nominative case marker *?e* and the oblique case marker *da* have the semantic feature of “[+vis]”. This indicates that both the *ino* ‘mother’ and *oko* ‘child’ were visible to the speaker when he made this utterance. Contrary to (135a), the nominative case marker *?o* and the oblique case marker *do* in (135b) carry the feature of “[−vis]”. This indicates that neither the mother nor the child were visible to the speaker when he made this utterance.

#### 4.9.2. Word Order

Most researchers into Tsou (Tung 1964; Zeitoun 1992; Chang 1998; Chang 2004) argue that the basic word order in Tsou is VOS (which is very different from the SVO order of Mandarin and Southern Min). To examine this claim, consider the examples in (136) and (137) cited from previous literature, and the example in (138) collected from my fieldwork. In (136), the subject *amo* ‘father’ is marked by the nominative case marker *?o*, and occurs in the sentence final position. The object *tatsəmə* ‘banana’ is marked by the oblique case marker *do*, and occurs before the subject. The verb *bonə* ‘eat’ appears before the object. Therefore, the word order of (136) is VOS. The same analysis can be applied to (137) and (138) whose word order is also VOS.

- (136) mo            bonə do tatsəmə ?o amo.  
tense2,Agent eat    Obl2 banana    Nom4 father  
‘My father ate a banana.’  
(Zeitoun 1992, p.15)

- (137) mi-ta            eobako da Mo?o ?e Basuya.  
tense4,Agent-3s hit        Obl1 Name Nom1 Name  
‘Bausya hits Mo?o.’  
(Chang 1998, p.15)

- (138) mo            ahdu bohtsinhi da Tsou ?e tsəmoi.  
tense2,Agent neg    attack    Obl1 people Nom1 bear  
‘The bear did not attack people.’

#### 4.9.3. Tense and Thematic Role Marking System

The tense and thematic role marking system is not directly related to the identification of the ML. However, most of the monolingual Tsou and bilingual Mandarin/Tsou examples provided in this thesis have tense and thematic role marking. It is, then, necessary to briefly discuss the functions of tense and thematic role marking system of Tsou.

As mentioned earlier, different linguists (Zeitoun 1992; Chang 1998; Chang 2004) have different views regarding tense in Tsou. Since Zeitoun's (1992) argument is more consistent, this thesis will adopt her perspective. However, because of the limited availability of Tsou data, I will still use Chang's (1998) examples to illustrate my points. Zeitoun argues that there are two types of tenses in Tsou, namely past and non-past, and proposes six tense markers in Tsou: *mi-*, *mio*, *mo(h)-*, *moso*, *i-*, and *o(h)-*. Each of them serves specific functions and contains different semantic information. These are illustrated in Table 3. These tense markers are also thematic role markers, which indicate the thematic role of the subject. The first four, namely *mi-*, *mio*, *mo(h)-*, and *moso* are used when the subject of a given sentence is an agent, while the remaining two, namely *i-* and *o(h)-*, are used when the subject is not an agent. This is also illustrated in Table 3. I will briefly describe the functions of each tense and thematic role marker in the following sections.



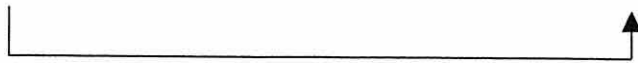
**Table 3: The Tense and Thematic Role Marking System in Tsou**

	Agent Marker		Non-Agent Marker
<b>Past tense marker</b>	<i>moso</i> (visible to the speaker)	<i>mo/moh-</i> (not visible to the speaker)	<i>o(h)-</i> (distant past)
<b>Gloss</b>	tense1, Agent	tense2, Agent	tense5, Non-Agent
<b>Non-past tense marker</b>	<i>mio</i> (free morpheme; marking on-going events /immediate past)	<i>mi-</i> (bound morpheme; marking on-going events /immediate past)	<i>i-</i> (present/recent past)
<b>Gloss</b>	tense3, Agent	tense4, Agent	tense6, Non-Agent

#### 4.9.3.1. Non-past Tense and Agent Markers: “mi-” and “mio”

The non-past tense marker *mi-* (marked as “tense4”) is a bound morpheme, which has to co-occur with other linguistic elements. It can either refer to “an event that is taking place at the time of speech”, as shown in (139a) or to “an event which has happened in the immediate past but whose result is still relevant to the time of speech” as shown in (139b) (Zeitoun 1992, p.46). The non-past tense marker *mi-* in (139a) indicates that the action *sæəsə* ‘wear’ is taking place when this sentence is produced. The *mi-* in (139b), however, shows that the speaker knows that Basuya just finished the action *mimo* ‘drink’, probably because of the alcoholic smell from Basuya’s body.

(139)a. **mi**-ta                    seoli'ano            s'æeəsə            **?e**            **Basuya** <sub>agent</sub>.

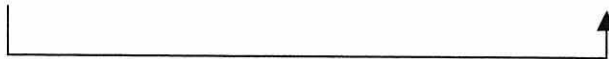


tense4,Agent-3s    just in time    wear    Nom1    Name

‘Basuya is wearing (something) now.’

(Chang 1998, p.43)

b. **mi**-ta                    mimo do            emi            **?e**            **Basuya** <sub>agent</sub>.



tense4,Agent-3s    drink    Obl2    wine    Nom1    Name

‘Basuya has drunk some wine.’

(Zeitoun 1992, p.47)

It is important to note that there is an agreement relationship between the tense marker and the thematic role of the subject. In (139b), the subject Basuya is the agent of the action “drink”. Hence, the agent tense marker *mi-* is used. The same analysis holds for another three agent tense markers mentioned below. Zeitoun (1992) argues that the tense marker *mio* (marked as tense3) is a free morpheme which only marks an event that is still going on at the time of the speech. This is illustrated in her example in (140).

(140) **mio**                    moiafo            **?e**            **amo** <sub>agent</sub>.



tense3,Agent    go out    Nom1    Father

‘Father is going out.’

(Zeitoun 1992, p.45)


Nether Zeitoun (1992) nor Chang (1998) specifies the difference between *mi-* and *mio*. From their discussion, it seems that both can refer to on-going events. The only difference shown in their studies appears to be that *mio* is a free morpheme and *mi-* is a bound morpheme.

#### 4.9.3.2. Past Tense and Agent Markers: “mo/mo(h)-” and “moso”

Both *mo/mo(h)-* and *moso* (marked as tense2 and tense1) are used to mark events which occurred in the past, but they have different semantic functions. According to Zeitoun (1992, p.48), “in a *mo* sentence, the event stated has been witnessed by the speaker while in a *moso* sentence it has not.” These points are illustrated by Chang’s examples (1998; 2004) in (141).

(141) **moh-ta**                    io    do                    Tfuya    **?e**                    Basuya

tense2,Agent,[+vis]-3s   live   Obl2[+loc]   Tfuya   Nom1,[+vis]. Basuya



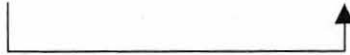
ne                    nətə?cə.

past                    last year

‘Basuya lived in Tfuya last year.’ (Chang 1998, p.43)

(142) **moso**                    tsmə?ho    **?o**                    Basuya.

tense1,Agent,[**-vis**]    come                    Nom4,[**-vis**]    Basuya



‘Basuya came.’ (Chang 2004, p. 16)

In (141), the nominative case marker *?e* and the past tense agent marker *moh-*, which have the information “[+vis]”, are used. This implies that the speaker not only knows



‘The wine has been drunk by the children.’

(Chang 1998, p.39)

The bold and underlined subjects in the three examples in (143) are patients of the main verbs. Hence, the non-agent markers *o-* and *i-* are used. Now, compare (143a) with (143b); both refer to an event that occurred in the past. However, in (143b) the *i-* indicates that the action “hit” has just happened, that is it had happened in the immediate past. The past tense marker *o-* in (143a) shows that the action “read” took place in a more distant past (e.g., one week or longer ago). (143c) shows that an event occurring in the present can also be indicated by *i-*.

#### 4.9.4. Summary

In this chapter, a brief description of the grammatical systems of Mandarin, Southern Min, and Tsou has been provided. This is necessary as one of the CS models this thesis aims to test, namely Myers-Scotton’s (2002a) MLF model, pays particular attention to the morphosyntax (i.e., word order and inflectional morphology) of the participating languages in code-switched utterances. For this reason, the discussion in this chapter focused on comparing the word order and inflectional morphology of Mandarin, Southern Min, and Tsou. Mandarin and Southern Min are typologically very similar languages. Both have the SVO word order and a limited amount of inflectional morphology. In contrast to Mandarin and Southern Min, Tsou has the VOS word order and an abundance of inflectional morphology. The case marking system in Tsou is particularly important because, according to the MLF model, case markers are the kind of inflectional morphemes (i.e., the outsider late system morpheme)(c.f. section 3.3.2), which can be used to identify the ML.

In addition to the basic word and inflectional morphology of Mandarin, Southern Min, and Tsou, this chapter has also discussed some language-specific elements. For instance, the functions of aspect markers, classifiers, and sentence final particles in Mandarin and Southern Min were described. Since these linguistic elements will be important in the discussion of the application of the MLF model to the Mandarin/Southern Min and the Mandarin/Tsou CS data, their similarities and differences were also presented. Finally, although the tense and thematic role marking system in Tsou is not directly relevant to the identification of the ML, they occur in most of the monolingual Tsou and bilingual Mandarin/Tsou examples provided in this thesis. It was then necessary to briefly discuss the functions of the tense and thematic role marking system in the Tsou language.

To sum up, the introduction of several linguistic elements in Mandarin, Southern Min, and Tsou in this chapter was a preparation for the application of the MLF model to the Mandarin/Southern Min and Mandarin/Tsou CS data in Chapter 6.

## Chapter 5: Methodology

### 5.1. Aims

This thesis has two major aims: (1) to test two morphosyntactic CS models, namely Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach with Mandarin/Tsou and Mandarin/Southern Min data. (2) to predict the future of the endangered language Tsou. To fulfil the second aim, it is necessary to investigate closely how Mandarin and Tsou are used in the conversations produced by the Tsou people from different generations. In this case, the sociolinguistic variable, namely age, becomes crucial.

Furthermore, according to the previous literature on endangered languages or language death (e.g., Wurm 1991; Crystal 2000), in many endangered language communities the language of the medium of education appears to influence people's language use. Since Mandarin is the language of medium of education in Taiwan, it is, then, necessary to take another sociolinguistic variable, namely educational level, into account. Hence, the subjects in the Mandarin/Tsou group will be divided into two sub-groups differentiated by their age and educational level. In order to make a comparison with the Mandarin/Tsou group, the subjects of the Mandarin/Southern Min group will be divided into three sub-groups according to the same sociolinguistic variables. Each subgroup will have ten subjects. From each subgroup, one hundred bilingual Mandarin/Tsou or Mandarin/Southern Min clauses produced by these subjects will be selected and analyzed in order to test the Myer-Scotton (2002a) and Muysken's (2000) models.

## 5.2. Educational Level and Age as Controlled Variables

For the Mandarin/Tsou speech community, two sample groups differentiated by age and educational level were selected:

Group 1: primary school education (or illiterate) / older Tsou people— 55 ~ 70

Group 2: secondary school education/ young Tsou people — 25 ~ 40

Because of Japanese colonisation (1895-1945) (c.f. Section 2.4.) and the implementation merely of a six-year compulsory education policy by the Taiwanese government in 1962, most of the older Tsou people received very limited education, mostly just to elementary school level, and some are illiterate (Chen 2001). After 1968, the government then extended the duration of the compulsory education to nine years (junior high school), which gave more opportunity for those people to receive education. However, because of their socioeconomic minority status, Tsou people cannot afford higher education. Even today, the great majority of Tsou people only receive secondary school education. In this context, one's age usually corresponds to one's educational level. Therefore, only two sample groups were selected from the Mandarin/Tsou speech community.

For the Mandarin/Southern Min speech community, the subjects were divided into three sub-groups with reference to their age and educational level:

Group 3: Educated below university level/older people— 50 ~ 60 years old

Group 4: University Level/older people — 50 ~ 60 years old



### Group 5: University Students — 18 ~ 25 years old

The subjects in Group 3 and Group 4 were born ten to fifteen years after the end of the Second World War, and experienced very harsh economic conditions with access to very limited educational resources (e.g., mass media). Only a small proportion of people in this generation, namely rich people and those who had achieved a very high standard in their high school education<sup>15</sup>, could enter university and continue their education. Thus, it is generally believed that those who did attend university and received their higher education through the medium of Mandarin speak Mandarin more frequently while those whose educational levels are low tend to use more Southern Min. Hence, it is hypothesised that the patterns of CS utterances produced by Group 3 and Group 4 will be different. To make a contrast with those in Group 4, the subjects from Group 5 (current university students) were selected in order to investigate whether there is a relationship between age and codeswitched utterances.

### 5.3. Organisation

The plan of this chapter is as follows: first for the Mandarin/Tsou speech community, a brief description of the locations where the data recordings were conducted will be provided. Second, the characteristics of Group 1 and Group 2 will be described. In relation to each group, background information, the reason for selecting the subjects, and the procedures for data recording will be given. Problems encountered and their solution will also be discussed. The same organization will be used in the description of the three subgroups in the Mandarin/Southern Min speech community.

---

<sup>15</sup> The educational system in Taiwan is different from that of the UK; it comprises elementary school, junior high school and high school. The general age to study in an elementary school is 7-12 (6 years); junior high school is 13-15 (three years); high school is 16-18 (three years).

## 5.4. Mandarin/Tsou Community

### 5.4.1. Geographical Distribution of Tsou Villages

Tsou people mainly inhabit the Alishan region (c.f. Figure 6) in the central part of Taiwan. Tsou villages, namely Pnguu, Tapangu, Lalauya, Niae'ucna, Saviki, Sinvi and Cayamavana, are scattered in the flat lands (between 400 and 1750 metres) of the Alishan region and their distribution is shown in Figure 6 below. It takes 15- 30 minutes to drive from one village to another by car. According to the census conducted in March 2005 by the Council of Indigenous People (<http://others.apc.gov.tw/popu/9403/aprp5803.htm>), the total Tsou Population is 6,056 of which 3,398 live in this region, while the other half lives in other cities. This study, however focused just on Tsou people who live in the Alishan region.

Figure 6: The Map of Taiwan

(Lonely Planet: [http://www.lonelyplanet.com/mapshells/north\\_east\\_asia/taiwan/taiwan.htm](http://www.lonelyplanet.com/mapshells/north_east_asia/taiwan/taiwan.htm), accessed on April 20, 2005)



Figure 7: The Geographical Distribution of Tsou Villages

(Map Source: The Local Government of Alishan Region (<http://www.alishan.gov.tw/>) (Accessed on April 15, 2005))



#### 5.4.2. Linguistic Situation

The great majority of Tsou people are Christians. The True Jesus Church, Presbyterian Church, and the Roman Catholic Church are the three main churches in Tsou villages. Most of the priests are Tsou people, though some are western missionaries. The Tsou people did not have a writing system. Hence, in the past important knowledge (such as hunting, or agricultural skills, fighting tactics, and medicine) and traditional values (such as religious belief, and mythology) were preserved and passed to the younger generations in an oral form. After the modernisation and industrialisation of Taiwan, the life style of Tsou people had changed fundamentally. Many Tsou people are

educated in Mandarin and work in big cities. Those traditional values and knowledge (including the Tsou language), which are not compatible to the modern society, have gradually been lost. Some influential priests from the churches (having had a little linguistic training) established an organization called *the Tsou Language Workshop* (鄒語工作室), and started to promote the use of a Romanised phonemic system either to record their oral history or to pass on the Tsou language to the next generation. Today, although all courses in the elementary schools in Taiwan are taught through the medium of Mandarin, the government allows the elementary schools in the Tsou villages to offer a two-hour Tsou language lesson each week. A number of publications have been produced by this workshop and are used as the teaching materials in the elementary schools in all the villages (seven schools in total).

#### 5.4.3. Group 1: Older Tsou people (age 55 ~ 70)

##### 5.4.3.1. Location: Tapangu

A pilot study was conducted in 2003. The village I visited was Pnguu, as one of my high school classmates was born there. Through her I was able to arrange accommodation in the village for several weeks. However, my classmate worked in the second largest city, Kaohsiung, and only returned to Pnguu when she had extended time of or for special occasions like the Chinese New Year. During my visit, she was only in the village for three days. This meant that she was not available for long to introduce me to Tsou people. As an outsider, it was hard to make contact with them, especially as the inhabitants were scattered, and there was little public transportation. Hence, an alternative location to Pnguu was sought for conducting the research.

The field research was conducted in Tapangu village. Tapangu has long been a

cultural and political centre and has several hundred inhabitants. The original local government was in Tapangu (although it moved to Lalauya while the field research was being carried out); it has a post office, police station, library, telephone company and other government administrative agencies in which Chinese people (i.e., native speakers of Mandarin and Southern Min) were dominant. The number of inhabitants in Tapangu was far greater than in Pnguu, and they were more concentrated in one area. The Tsou people here are more open to outsiders. These factors influenced the selection of Tapangu rather than Pnguu as the location for collecting natural speech uttered by Tsou people.

#### 5.4.3.2. Subjects and Settings

As a complete outsider to the Mandarin/Tsou speech community, I did not speak Tsou, and had very little knowledge of the current situation of the Tapangu village from previous studies related to Tsou. From the publications edited by the *Tsou Language Workshop*, I acquired contact information for Mr. Wang Xingshi, one of the editors who lived in Tapangu. In order to have better understanding of this village, I made an appointment with Mr. Wang before my first trip to Tapangu in mid-August 2004. After several hours of discussion with him, I become clearer about the daily routine of Tsou people, such as the usual times and places of their social gatherings. Moreover, with his assistance several possible candidates for conducting the data recording were selected. On the day following this meeting, I started to visit these candidates in person and asked whether they would participate in this research.

After observation lasting for several days, I was fortunate to find a kind woman (referred to below as “Assistant 1”) who was willing to do the recording. Assistant 1 owned a small grocery store and a very small restaurant (with just two tables) in

Tapangu village. During the day, Tsou people usually work on the farms outside their village, so the village was almost empty. In the evening, many people would sit around the fire and have a barbecue or drinks together. Some of them would go to Assistant 1's little restaurant and have their dinner with some drinks. As the restaurant was very small the number of customers was limited. Hence, it was a better environment for recording conversations than the setting where a group of people participated in a barbecue, drinking and singing. Ten subjects (including Assistant 1) were selected. Assistant 1 was given detailed information about the recording plan, including the proper age range of the potential subjects, and the duration of the recording plan (about one month, though this was extended to nearly two and a half months because of various practical difficulties which are specified in the following section).

The other nine subjects were established customers of this restaurant (eight males and one female), and were selected by Assistant 1. Their age was between fifty-five and seventy years old. They were all Mandarin-Tsou bilinguals. Some of them, especially those who were above 65 years old, were also fluent in Japanese.

Although these ten subjects were acquaintances of each other, it was hard to gather them together. This is because some of them lived very far away from the Tapangu village. However, they had frequent contact with Assistant 1, and went to restaurant separately every two or three days. Assistant 1 asked the subjects individually if they would participate in this research. After acquiring oral permission from the selected subjects, Assistant 1 was asked to be sure to do several things before any recording was conducted:

1. That all subjects were clear about the recording plan, namely that until they were informed that the plan was complete, their conversations in any informal settings would be recorded without any further notice to them.
2. That all subjects had completed the questionnaires and signed the letters of consent.
3. That all subjects had been informed that they were free to withdraw from this research if they changed their mind later.

Since some of the selected subjects were illiterate, Assistant 1 had to answer the questionnaires for them and sign the letters of consent on their behalf. Some of them were willing to sign the letters of consent, but refused to provide their personal details except their age. Hence, the information they provided was not so complete as in other groups.

#### 5.4.3.3. Data Recording

All the recordings were made in the restaurant. The digital recorder was placed on the counter, which was five metres from the tables where the subjects sat. The recorder was put behind a vase in order to avoid the subjects being too aware of it. Originally, I planned to spend one month collecting data from this group, from mid-August to mid-September 2004. However, because of serious destruction of the roads and floods caused by several typhoons from August to October, I had to finish some of the recordings at the end of October. Whenever the weather permitted, I would go to Tapangu on a Thursday, and would stay there for one night due to my limited budget. Every Friday, I collected the recordings from Assistant 1 and transferred them to my laptop. Since Assistant 1 did not ask for payment, I usually took gifts (e.g., fruit, cakes and etc.) to her in order to express my gratitude.



#### 5.4.3.4. Data

##### *Finding A Translator*

Since I did not speak the Tsou language, I looked for an appropriate person to transcribe the data and translate the transcription into Chinese during my stay in Tapangu. Such a person had to be fluent in both languages, and to have a good command of the Chinese writing system as well as to be able to transcribe the recordings using a Romanized system. The priests were the most suitable candidates, but they were too busy to do this task. The Tsou speakers in Tapangu were mainly older people who were not familiar with Chinese writing. Others, who were able to do the task, seemed to show little interest, though they were informed that a payment of £100 or more would be offered. Finally, I found a young man – Basuya<sup>16</sup> from the Cayamavana village who came to Tapangu to visit his relatives. Basuya was 28 years old and used to be a soldier. When he was a junior high school student he had learned the Romanized system, which was developed specifically for Tsou, from a priest, and was familiar with the Chinese writing system. He was also a fluent speaker of Tsou. I briefly introduced him my research plan, namely to record the conversations of two age groups (elders and younger people) for linguistic analysis, and he agreed to assist in this research.

##### *Data Transcription*

After collecting the recordings from Assistant 1 and saving them on my laptop, I then went to the Cayamavana and gave Basuya the recordings on the next day. The total length of the recordings was about nine hours. Basuya usually transcribed the data by using the computer in the True Jesus Church or my laptop. I discussed the translation and gloss with Basuya during each of my stays in the village. In the recordings, any

---

<sup>16</sup> My translator, Basuya e Yakumangana, requested that I give his name in this research.

utterances in Mandarin were transcribed using the Chinese writing system, and those in Tsou were transcribed using a Romanized system with detailed gloss. All the Tsou words were then translated into Chinese.

Tsou is a language with abundant inflection morphology. Ideally, it would have been best to translate every morpheme rather than every word. However, this would have been a very difficult task for a person like Basuya who had no linguistic background. He was only able to translate word by word, and could not go further into the grammatical function of each morpheme. In fact, I consulted some priests, members of the Tsou Language Workshop, but they were not able to solve my problems. However, previous research on Tsou provided detailed information about Tsou morphology and negation (Chang 1998; Ko 2001; Zeitoun 1992). Following the information in those studies, I attempted to analyze the meaning of each morpheme myself and then, each time we met, checked this with Basuya and Mr. Zhen Xin-de, who was a priest of the True Jesus Church in the Cayamavana village and a member of the Tsou Language Workshop.

Basuya raised some additional problems when transcribing the recordings (e.g., background information to the conversation and recognizing the speakers). I sought help from Assistant 1 with those problems when I went to Tapangu. The gloss, translation, and transcription were checked by Mr. Zhen Xinde. Finally, the first one hundred bilingual clauses were extracted from the transcription and translated into English.

#### 5.4.4. Group 2: Younger Tsou People (age 25 – 40)

Finding the appropriate target subjects of this group was not as easy as for the

previous one. Because of the strict language policy in the past and the current socio-economic demands in Taiwan, young Tsou people's proficiency in their native language, namely Tsou, seems to have sharply declined. I found that people under twenty-five seemed to only use Mandarin in daily conversation.

In Tapangu, about five minutes walk from the restaurant where the recordings of the older group were conducted there was a café. During the day, the customers were mainly those who worked in the government sectors. In the evening, many young Tsou people (from teenagers to those aged about thirty) would gather in the café after school or work because they could read a variety of Chinese popular comic books, novels, or magazines while they were having their meals or drinks. In order to find potential subjects for Group 2 (the younger Tsou group), I normally had dinner in that café on each of my stays in Tapangu. Each time I spent more than two hours there to record their conversations (around thirty minutes) and to observe which language they used. They were not informed before or after the recordings were conducted, as the recorded data were only trials and were not to be used for analysis. The results showed that they all spoke Mandarin, and almost no Tsou was used.

I also visited other villages like Lalauya, Niae'ucna, and Sinviki, and adopted the same strategy; the results were similar. In each village, two young people (aged between twenty to twenty-five) were also informally interviewed and were asked: "What languages do you usually use when talking to your friends or your family members?" Both of their answers were Mandarin, which corresponded to what I had observed. Such a situation posed some difficulties for finding appropriate subjects. My translator Basuya said that young people in Cayamavana village might speak more Tsou, for that village had had much less contact with Chinese people compared

with Tapangu.

#### 5.4.4.1. Location : Cayamavana Village

The population of Cayamavana is over four hundred of which about 60% is Tsou; 30% is Bunun, a different aboriginal tribe (speaking a different language); only 10% is Chinese. Compared with Tapangu many fewer tourists visit this village and then mainly during holidays or at weekends. The major industry here is agriculture and the products include sweet potatoes and betel nuts. Hunting is the other important economic resource. Recently, because of government's policy to encourage tourism, some small B & B's owned by Tsou people have been opened in order to attract more tourists.

#### 5.4.4.2. Subjects

In the initial stage of searching for subjects a great deal of observation was conducted. In addition to the restaurants, I also tried places such as grocery stores and elementary schools. The results resembled the observation in the other villages, Mandarin was the only language used by young people.

At the end of December 2004, the manager of a recently developed tourist spot, Chayi Farm (about one hour's driving from this village), planned to hire some Tsou people to perform traditional dancing and singing at the farm. A performance group which included more than ten Tsou people aged from twenty-five to forty and included my translator, signed a contract with the manager to present their show for a season. Every evening the members of this performance group would sit around the fire, and discuss or rehearse what was to be presented on the next day.

The leader of this group was Basuya's (my translator) brother, and all the members were his friends. All group members were bilinguals in Mandarin and Tsou. Basuya introduced me all the group members and invited me to one of their rehearsals. Before actually choosing them as subjects, I observed the languages they used. The situation was better than in the villages. Mandarin was again the major medium of their conversations; some switches between the two languages occurred, though the frequency was not high. After getting to know them better, I briefly introduced the plan to record their conversations. They all kindly agreed to participate in the research. Six males and four females were selected as the subjects. They then completed the questionnaires and signed the letters of consent. Similarly, they were fully aware that their conversations would be recorded without any further notice. They knew that the research was related to linguistics, but the specific focus was unknown to them.

#### 5.4.4.3. Data Recording

The recording task of Group 1 (older Tsou people) was completed at the end of October 2004. Almost two months were spent for searching for appropriate subjects for Group 2 (younger Tsou people). The recordings of Group 2 started at the beginning of December 2004 and continued to the end of February 2005. One digital recorder was given to Basuya, and he recorded the subjects' conversations in rehearsals or on other occasions where some of the subjects drank wine and chatted together. The content of the recorded conversations was mainly about their rehearsals and shows in Chayi Farm, plans to organize activities in all the Tsou villages, and discussion of the future of Tsou people. The subjects were given the week to do the recordings, and I went to Cayamavana each weekend to collect the recordings. Unfortunately, most of the conversations were in Mandarin and, ideally more recordings should have been done. However, because of limited amount of time and

budget, the fieldwork had to stop.

#### 5.4.4.4. Data

##### *Translator*

During the process of recording data in Cayamavana, a young woman, Danivu<sup>17</sup> from Sinvi, another Tsou village, was introduced to me by Basuya. She agreed to participate in this research as a translator. Danivu was a university student and majored in ethnographical studies. Although her proficiency in the Tsou language, similar to other young people, was not good, her parents and relatives were all fluent speakers. Hence, she could consult them on encountering problems of comprehension. She was paid after completing the transcription and translation.

##### *Data Transcription*

The total length of the recordings was about six hours. Most of the conversations were in Mandarin, and the quality of some recordings was very poor. Therefore, only forty-one minutes could be used for analysis. First, the recordings were transcribed using the Chinese writing system and a special Romanized system designed for Tsou, and then translated into Chinese with detailed glosses. The same procedures described in section 5.4.3.4. were followed to analyse the grammatical information carried by each morpheme in Tsou. The transcription and translation were then double-checked by Mr. Zhen Xin-de. Only thirty bilingual Mandarin/Tsou clauses were found in these recordings, and they were extracted from the transcription and translated into English.

### 5.5. Mandarin/Southern Min Community

#### 5.5.1. Tainan Area

---

<sup>17</sup> My second translator, Danivu e Tapangu also requested that I give her name in this research.

Tainan City is located on the southwest coast of Taiwan (c.f. Figure 6). With a population of 755,536, it is now the fourth largest city (Tainan City Council: <http://www.tncc.gov.tw/>; accessed on April 10, 2005). Between three hundred and four hundred years ago, settlers from the southern part of Fujian Province (also known as Min) of China (c.f. Figure 1 repeated as Figure 8 below) went to Taiwan to seek more farm lands, and built their first settlement in Tainan area. These settlers were known as the Southern Min, and the language they spoke was South Min, which is also called Taiwanese nowadays.

Figure 8: Taiwan Strait Area

(Map Source: The University of Texas, Austin: [http://www.lib.utexas.edu/maps/cia04/taiwan\\_sm04.gif](http://www.lib.utexas.edu/maps/cia04/taiwan_sm04.gif)

Access Date: April 10, 2005)

Taiwan Strait Area



802566 (R01492) 8-98



Because of its long history, the Tainan area (including Tainan City and County<sup>18</sup>) is usually seen as the place where most of the traditional Taiwanese customs are preserved. Since the Japanese colonial period, Tainan county has been the agricultural centre and Tainan City has been one of the light industry centres in southern Taiwan. Thus, the majority of the residents in greater Tainan area are either farmers or factory workers; they are the major forces supporting the political party DPP (the Democratic Development Party). The main political aims of the DPP are to oppose the regime of Mandarin Chinese speakers established by the Chinese Nationalist Party (c.f. section 2.5), and to pursue the political independence of Taiwan. Because of political ideology and these historical factors, Southern Min tends to be used more frequently in the Tainan area compared with other cities in the central and northern parts of Taiwan. For this reason, Tainan City and Tainan County were chosen as the locations for collecting the Mandarin/Southern Min CS data. The other reason why the Tainan area was chosen was that it was my birthplace. Hence, the social networks that might be useful for collecting the target data were readily available.

### 5.5.2. Group 3: Educated below University Level/Older People (age 50 ~ 60)

#### 5.5.2.1. Subjects and Setting

Ten people, including three males and seven females, from the Environmental Protection Department of Yong Kan City in Tainan county were selected as the subjects for this group. Five of them had graduated from high schools, and the other five had graduated from junior colleges (an educational level between high school and university). All were fluent speakers of both Mandarin and Southern Min. They had all worked in the department for many years, and were close friends. The reason for

---

<sup>18</sup> Tainan County is the area around Tainan City.

selecting these subjects was that one of my relatives (referred to as “Assistant 2” below), was one of the selected subjects and worked in this department, hence, it was relatively easy to approach these subjects.

The main function of the Environmental Protection Department is to oversee all the environmental maintenance works of the city (e.g., collecting rubbish, and administration related to the protection of the environment). Staff work either in the office or in the actual collection of rubbish. Except when they answered phone calls from the general public or attended regular meetings held by the county council, these subjects seldom had contact with people from outside the group (e.g., from other departments or the general public) during working time. They often engaged in informal conversation among themselves on topics including TV drama, news, or events in their everyday lives.

#### 5.5.2.2. Meeting the Subjects

As Assistant 2 was already a member of the target group, with her help I arranged a meeting with another nine subjects and explained to them that their conversations would be recorded and then analyzed for research purposes. They were also told that in the research their conversation would be analyzed anonymously. These ten subjects (including Assistant 2) were only informed that they would be involved in a linguistic research project, but the specific aim of the research, i.e. examining the morphosyntactic structure of Mandarin/Southern Min codeswitched utterances, was not disclosed to them. During the meeting, ten questionnaires with letters of consent attached were distributed to the subjects (c.f. Appendix 1 and 2). In the questionnaires, the subjects were asked to provide background information (i.e., name, age, educational background, contact details and so on), and to answer seven questions

about the languages they used for communication in daily life. The subjects were then asked to sign the letters of consent in order to provide permission for recording their conversations without further notice. In addition, they were informed that they would be free to withdraw from this research later without giving reasons.

#### 5.5.2.3. Data Recording

The aim of the data recording was to record the naturally occurring conversations of the selected subjects. To eliminate the effect of the “observer’s paradox” proposed by Labov (1972, p.209), I was not involved in the data recording process, which was conducted in the office of the Environmental Protection Department during office hours (8:00 a.m. – 17:30 p.m.). Assistant 2, who was already a member of the target group, did the recording using an ENOX digital recorder whose size was 9.5 cm x 3.5 cm x 1.5 cm. The recorder was put on the desk and covered by different objects (e.g., a purse, a piece of paper, or folders), and thus did not attract the subjects’ attention during the recording. To ensure the quality of the recording and for the convenience of data transcription, Assistant 2 was told to start the recording only on occasions when 4-5 people were in the office, and that the length of each conversation being recorded should be greater than 5 minutes. The duration of the overall recording process was one month.

#### 5.5.2.4. Transcribing Data

Assistant 2 met me every weekend and gave me what she had recorded for the whole of the preceding week. All the recorded data were then transferred onto the laptop immediately. The recordings were then transcribed. Any problems encountered, such as identifying the voices of individual speaker, were discussed in the next meeting.

The length of the recordings was about 9 hours and 15 minutes. Telephone conversations in the recordings were excluded when transcribing the data because the usual length of such conversations was only 1-2 minutes including some long formulaic greetings, and the actual parts that could be used for analysis were very limited. Telephone use would probably make their speech more formal, and, therefore, it was excluded.

#### 5.5.2.5. Transcribing Procedures

The main objective was to obtain the first one hundred Mandarin/Southern Min bilingual clauses from the transcription. Once the target number of switches was met, the work of transcribing recordings was stopped. In the transcription, any utterances in Mandarin were first transcribed using the Chinese writing system. Then, any Southern Min expressions were transcribed using a phonemic system based on the IPA (International Phonetic Alphabet) system because of the lack of a specific phonemic writing system for Southern Min. Second, the Chinese version of the transcription was transferred to the Romanized Pin-Yin system, which was designed specifically for Mandarin, and the Southern Min part was represented by the IPA system. Third, the bilingual Mandarin/Southern Min clauses were then extracted from the transcription, and a detailed linguistic gloss of each morpheme was provided. Finally, the all the extracted bilingual clauses were translated into English.

#### 5.5.3. Group 4: University Level/Older People (age 50 ~ 60)

##### 5.5.3.1. Subjects and Settings

There were ten subjects, seven females and three males, in this group. They were either current or retired teachers from a junior high school (pupils' age: 13-15) in

Tainan county (just outside Taiwan City), and all had received university education. All of the selected subjects were fluent speakers of Mandarin and Southern Min, and their ages ranged from 50-60 years. Although some of the subjects were retired, they still had frequent contact with each other and with those colleagues who were still teaching in school. Every week, they would meet several times in a café, and sometimes they would go away together for a short trip of around 1-2 days. My mother was a former teacher of this school, and was the contact person for the selected subjects. Since my mother did not join in their activities very often, one of her former colleagues (a current teacher), referred to as “Assistant 3” below, was selected to do the recordings. The recordings were conducted in two different locations in Tainan City (within Tainan county). About half of the eleven hours recording took place in the café while the other half was made in the car on a two-day trip. The topics of their conversations were mainly families, colleagues, or school affairs.

#### 5.5.3.2. Meeting the Subjects

With my mother’s assistance, I attended one of the subjects’ meetings in a café. During the meeting, the recording plan was explained to them and then the volunteers who were willing to participate in the research project were chosen. Since only ten subjects were needed for each group, ten out of fifteen people attending the meeting were selected on the basis of who arrived first. Then, these ten were informed that their conversations in the later meetings would be recorded without any further notice and the recordings would be analyzed anonymously for research purposes. The same questionnaires and letters of consent as in the previous group, were distributed to the subjects. In the questionnaires, they (including Assistant 3) were only informed that they would participate in a linguistic research project, but the actual research aim was

not specified.

#### 5.5.3.3. Data Recording

The duration of the entire recording process of this group was one month. The total length of the recordings was about eleven hours, of which six hours took place in a café in Tainan city and the five hours were recorded in two cars during a two-day trip to Ken Ding National Park in southern Taiwan. Assistant 3 was responsible for the data-recording task. Seven selected subjects joined this trip. In order to record the conversations of all subjects during the trip, my assistant sat in one car on the way to the national park and changed to the other on the way home. The digital recorder was placed on Assistant 3's lap and slightly covered by her hat, so that the subjects did not know their conversations were being recorded, but they were informed at the end of the trip. For the recording in the café, the digital recorder was placed on the table and covered by Assistant 3's handbag. To ensure the quality of the recording, Assistant 3 only conducted the recordings on occasions when only five or six subjects met together.

#### 5.5.3.4. Transcribing Data

Since my mother was acquainted with the selected subjects, she was consulted if there were any problems with identifying the speakers or if any background information was needed. Assistant 3 met me every other week and gave me the recorded data and also helped me to solve the problems encountered. Then, the recordings were transcribed following the same procedures outlined in section 5.5.2.5.

### 5.5.4. Group 5: University Students (age 18-25)

#### 5.5.4.1. Subjects

Ten students, including three males and seven females, of the Chia-Nan University of Pharmacy and Science in Tainan County were selected as the subjects for this group. My father is an associate professor of this university. He hired one student whom he knew well as the assistant (referred to below as “Assistant 4”) to conduct the recordings. Assistant 4 was paid on a monthly basis. The other nine subjects were Assistant 4’s classmates, flat mates or friends who studied in the same university. All the subjects were fluent speakers of Mandarin and Southern Min.

#### 5.5.4.2. Meeting the Subjects

I first made an appointment with Assistant 4 and explained to her the plan of recording the conversations between her and her friends. As in the case of the other assistants, Assistant 4 did not know the main focus of my research. She was asked to find another nine subjects aged between 18-25 using her own social network. Then, ten copies of the questionnaire and letter of consent were given to her. In contrast to the other two groups, some of the subjects selected did not know each other. It was difficult to arrange a meeting for all of them. Therefore, Assistant 4 explained the details of the recording plan to them individually, and she was asked to make sure that several things were done before any recording was undertaken:

1. That all subjects were clear about the recording plan, namely until informed that the plan was completed, their conversations in any informal settings would be recorded without notice.
2. That all subjects had completed the questionnaires and signed the letters of consent.
3. That all subjects had been informed that they were free to withdraw from this research if they changed their minds later.

#### 5.5.4.3. Data Recording

The duration of the overall recording process took three months, starting at the beginning of October and ending at the end of December 2004. The total length of the recordings was about 28.5 hours. All the recordings were conducted by Assistant 4. The locations where the recordings were made included the campus (e.g., car parks, or lecture rooms), Assistant 4 and her classmates' flat, shopping centers, some tourist spots they visited or other informal settings. The digital recorder was always covered by some object in order to hide it from the subjects' attention. The topics of conversation included university studies, clothes, fashion accessories etc.

#### 5.5.4.4. Transcribing Data

Assistant 4 met me every other week to discuss the problems encountered when transcribing the data and gave me the new recordings. The recordings were transcribed following the procedures outlined in section 5.5.2.5.



## Chapter 6: Data Analysis

### 6.1. Introduction

One of the research questions that this thesis attempts to answer is whether there is a universally applicable CS model? Myers-Scotton (2002a; 2006) argues that the MLF model has universal applicability. Muysken (2000), however, argues that the MLF model only accounts for a single CS pattern, namely insertion. To give a fuller account to CS, one must also consider the other two CS patterns i.e., alternation and congruent lexicalization. In Chapter 3, the main arguments of Myers-Scotton's (2002a) MLF Model and Muysken's (2000) typological approach were briefly reviewed and evaluated with reference to published data. This chapter will continue the evaluation of these two models and test them using the Mandarin/Tsou and Mandarin/Southern Min CS data.

According to Myers-Scotton (2002a, p.8), "the unit of analysis of the MLF model is a bilingual CP." To identify the ML of a bilingual CP, we need to check both the word order (i.e., the morpheme order principle) and one type of inflectional morpheme (i.e., the system morpheme principle) (c.f. section 3.3). These two key principles of the MLF model have been applied to different sets of CS data, such as Swahili/English (Myers-Scotton 1993/1997); Zulu/English & Sotho/English (Finlayson, Calteaux, & Myers-Scotton 1998); German/English (Fuller & Lehnert 2000); Mandarin/English (Wei 2001) and Welsh/English (Deuchar 2006). The premise of work to date applying the MLF model is that the participating language pairs should have different word orders or different inflectional morphology. This raises a potential problem: does the MLF model apply to language pairs which have exactly the same word order and have very limited inflectional morphology, as in the Mandarin/Southern Min data? In this chapter, the MLF model will be tested with Mandarin/Southern Min data in order to

find the answer to this question.

In comparison with the MLF model, Muysken's (2000) typology approach deals with a wider range of data. The MLF model looks only at intra-sentential switches (i.e., switches occurring within one sentence) and uses the CP as the basic unit of analysis, while the typology approach not only deals with intra-sentential switches, but also looks at inter-sentential switches (i.e., switches occurring between sentences) and switches that occur between different turns. In this chapter, I shall consider whether Muysken's model is more suitable for my data than the MLF model. However, a basic problem of Muysken's (2000) model was found when applying it to the Mandarin/Southern Min and Mandarin/Tsou data, namely there is a lack of a basic unit of analysis. Hence, the criterion for defining a switch is rather unclear. To solve this problem, I will adopt a clause-based approach proposed by Deuchar, Muysken and Wang (2007) and discuss the theoretical and empirical problems of Muysken's (2000) model.

The organization of this chapter is as follows. First, I will test Myers-Scotton's (2002a) MLF model with the Mandarin/Tsou and Mandarin/Southern Min data, and propose some revisions to the model. Second, I will test Muysken's (2000) typological approach with the same sets of data. Similarly, I will address the theoretical problems encountered (i.e., how to define a switch) and discuss some possible modifications of the original model. Finally, I will evaluate the revised MLF model and Muysken's typological approach and discuss which model is more successful in terms of describing the Mandarin/Tsou and Mandarin/Southern Min data.

## 6.2. Testing the MLF Model

This section aims to test the MLF model, which includes two important principles, namely the morpheme order principle and the system morpheme principle. For the convenience of analysis, I will repeat the definitions of these two principles, which were introduced in section 3.3.1.2.

### “The Morpheme Order Principle:

In ML + EL constituents consisting of singly-occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.

### The System Morpheme Principle:

In ML + EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e., which participate in the sentence’s thematic role grid) will come from the ML.”

(Myers-Scotton 2002a, p.59)

According to these two principles, if the ML of a bilingual clause is language A, the surface word order of the entire clause will be that of language A, and the outsider late system morphemes (e.g., subject-verb agreement affixes or case markers) will also be provided by language A. In the following discussion, these two principles will be tested with the Mandarin/Tsou, Mandarin/Southern Min data. Some theoretical problems caused by the linguistic nature of these three languages, especially Mandarin and Southern Min, will be raised and discussed.

### 6.2.1. Mandarin/Tsou Data

To recall the discussion in Chapter 3, Mandarin is a SVO language, while Tsou is a VOS language. Unlike Mandarin, a language in which inflectional morphology is almost absent, Tsou has a very complex morphological system. In the following sections, we shall test the two important principles in the MLF model by using our bilingual Mandarin/Tsou data.

#### 6.2.1.1. Method of Analysis

The aim of this section is to show how the MLF model is to be applied to analyse our Mandarin/Tsou bilingual data. (144a) is a bilingual clause collected from one of my Mandarin/Tsou data sets. From the analysis of the examples in (144), I will illustrate how the matrix language is identified. The monolingual Mandarin and Tsou versions are provided in (144b) and (144c).

(144)a. mo                    *gan-kwai* *ma-shang* *pao-diao* ?o Basuya subject  
tense2,Agent    quickly    right-away    run away    Nom4    Name  
'Basuya ran away quickly.'  
(Mandarin/Tsou; normal print = Tsou; *Italic* =Mandarin)

b. Basuya subject    *gan-kwai*    *ma-shang*    *pao-diao*. (Mandarin)  
Name                    quickly        right-away    run-away  
'Basuya ran away quickly.'

c. mo                    aupətsio    p?kako    ?o Basuya subject. (Tsou)  
tense2, Agent    quickly    run away    Nom4    Name  
'Basuya ran away quickly.'

If we compare the three examples in (144), it is clear that the bilingual clause in (144a) has the same word order as the monolingual Tsou example in (144c), for the subject

occurs at the sentence final position. According to the morpheme order principle, the surface word order of a bilingual clause must come from the ML. We may say that Tsou is likely to be the ML. However, to identify the ML the second criterion, namely the system morpheme principle, also needs to be checked. If the ML is Tsou, as the morpheme order principle predicts, then, the outsider late system morpheme must also come from Tsou. To recall the discussion of the 4-M model in section 3.3.2, there are four types of morphemes, namely content morphemes, early system morphemes, and late system morphemes. The late system morphemes include bridge morphemes and outsider late system morphemes. Myers-Scotton's (2002a, p.75) definition of these four types of morphemes is cited below:

“Content morpheme: a morpheme which assigns or receives thematic roles

e.g., all nouns and verbs’

Early system morpheme: a morpheme which is indirectly elected by its head

content morphemes’

e.g., determiners and plural markers

Bridge morpheme: a morpheme which connects content morphemes with

each other without reference to the properties of a head

e.g., possessive ‘of’ and ‘s.’

Outsider late system morpheme: a morpheme which looks outside its own

immediate maximal projection for

information about its form.’

e.g., subject-verb agreement or case affixes”

According to the system morpheme principle, only the system morphemes, which “have grammatical relations external to their head constituent” must come from the matrix language (p.87). In other words, if we want to identify what the ML of a bilingual clause is, it is necessary to check which language provides the outsider late system morphemes. Because of the limited number of research studies in Tsou grammar, the details of Tsou’s tense and case marking system are still not very clear. However, some clues can be found from current literature. Now, consider (144a), which is repeated as (145) below.

- (145) mo                    *gan-kwai* *ma-shang* *pao-diao*    ?o        Basuya.  
                                  tense2,Agent quickly    right-away    run away    Nom4    Name  
                                  ‘Basuya ran away quickly.’  
                                  (Mandarin/Tsou; normal print = Tsou; *Italic* =Mandarin)

According to the system morpheme principle, case affixes are outsider late system morphemes. In (145), one outsider late system morpheme is found, namely the Tsou nominative case marker “?o”. The occurrence of the Tsou case marker indicates that Tsou is the ML. Since the results of morpheme order principle and the system morpheme principle both point in the same direction, we can conclude that Tsou is the ML of (145).

One more Mandarin/Tsou bilingual example with a different ML is provided in (146a). Again, the monolingual Mandarin and Tsou versions of (146a) are provided in (146b) and (146c).

- (146) a. homo *bai-tian wen bu-dao zhu-she.* (Mandarin/Tsou)  
 If daytime smell neg.-to piggery  
 ‘If at daytime, you can’t smell the piggery.’  
 (normal print = Tsou; *Italic* = Mandarin)
- b. *ru-guo (zai) bai-tian wen bu-dao zhu-she.* (Mandarin)  
 If (at) daytime smell neg.-to piggery  
 ‘If at daytime, you can’t smell the piggery.’
- c. homo hie ah?ta eləi ?o emo no fəeə. (Tsou)  
 if daytime neg. smell nom4 home poss. pig  
 ‘If at daytime, you can’t smell the piggery.’

The bilingual clause in (146a) has the same word order as the monolingual Mandarin sentence in (146b). Note that in Mandarin the negation marker “*bu*” occurs after the main verb *wen* ‘smell’, while in Tsou the negation marker “*ah?ta*” appears before the main verb *eləi* ‘smell’. According to the morpheme order principle, Mandarin is the most likely candidate as the ML. To identify the ML of a bilingual CP, the other criterion also needs to be taken into account, namely the system morpheme principle. If the ML of (146a) were Tsou, then we would expect the appearance of outsider late system morphemes (i.e., the nominative case marker *?o*), as shown in (146c). However, no such affixation is found in (146a), so, this points to Mandarin as the ML. Since both principles have the same result, it is concluded that Mandarin is the ML of (146a).

#### 6.2.1.2. Results

Two sets of Mandarin/Tsou data, which were collected from Group 1 (older Tsou people aged 55 - 70) and Group 2 (younger Tsou people aged 25 - 40) (c.f. section

5.4.3 and 5.4.4), were used for the test of the MLF model. One hundred bilingual clauses were collected from Group 1. Since Mandarin seems to have replaced Tsou as the major language used by young Tsou people, in around six-hours of recording I was only able to collect thirty bilingual clauses from Group 2. The method discussed in the previous section was adopted to identify the ML of each bilingual clause collected from Groups 1 and 2. The results are shown in Table 4.

**Table 4: Results of the Application of the MLF Model to Mandarin/Tsou Data**

	<b>Mandarin = ML</b>	<b>Tsou = ML</b>	<b>Total Number of Bilingual CPs</b>
<b>Group 1</b>	21 (21%)	79 (79%)	100
<b>Group 2</b>	20 (67%)	10 (33%)	30

As Table 4 shows, the ML of 79% of the data collected from Group 1 is Tsou and 21% is Mandarin. For Group 2, the ML of 67% of the data is Mandarin and 33% is Tsou. Hence, the MLF model was successfully applied to the two Mandarin/Tsou data sets. No problematic data (i.e., data whose ML was unable to be identified) were found. Since the aim of this section is only to test the applicability of the MLF model, the results and their implications will not be discussed here. A detailed analysis of these results and the problematic data will be provided in Chapter 7: Re-analysis of the Data Using the Revised MLF Model.

## 6.2.2. Mandarin/Southern Min Data

### 6.2.2.1. Theoretical Problems of the MLF Model

In Chapter 4, I showed that the major difference between Mandarin and Southern Min is at the lexical and phonological level, and their syntactic structures are almost the same (both of them have SVO order). Hence, whether the morpheme order principle



can be applied to identify the ML of a Mandarin/Southern Min bilingual clause is questionable. (147a) is a bilingual clause collected from one of the Mandarin/Southern Min data sets. The monolingual versions in Mandarin and Southern Min of (147a) are also provided in (147b) and (147c).

(147) a. **wa** *shi-yi-dian* *shui*. (Mandarin/Southern Min)

I eleven o'clock sleep

'I went to bed at eleven o'clock.'

(**Bold**=Southern Min; *Italic*= Mandarin)

b. *wo* *shi-yi-dian* *shui*. (Mandarin---SVO)

I eleven o'clock sleep

'I went to bed at eleven o'clock'.

c. **wa** **zap-i-diam** **kun**. (Southern Min---SVO)

I eleven o'clock sleep

'I went to bed at eleven o'clock'.

According to the morpheme order principle, the surface word order of a bilingual clause has to be checked to identify the ML. However, as the two monolingual clauses in (147b) and (147c) suggest, Mandarin and Southern Min only have lexical and phonological differences, whereas the word orders are the same (i.e., SVO). Hence, whether the word order of (147a) belongs to Mandarin or Southern Min is unable to be identified. This suggests that the morpheme order principle is not applicable to the Mandarin/Southern Min data.

Furthermore, in the bilingual clause in (147a) no outsider late system morpheme is found (e.g., subject-verb agreement affixes or case markers). There is no indication of

any agreement relationship between the Southern Min subject *wa* ‘I’ and the Mandarin verb *shue* ‘sleep’. Even in the monolingual version in (147b) and (147c), there is no subject-verb agreement. Hence, the system morpheme principle is not applicable. Since neither the morpheme order principle nor the system morpheme principle is applicable, the ML of (147a) is unable to be identified.

(148a) is another example from my Mandarin/Southern Min corpus. Similarly, the monolingual versions in Mandarin and Southern Min are provided for comparison.

(148) a. **ze** **dwe** **dʒi-lei** **ʃi** *xiang-cai?* (Mandarin/Southern Min)

This which one-class cop coriander

‘Which one of these is coriander (flavour)?’

(**Bold**=Southern Min; *Italic*=Mandarin)

b. *zhe na yi-ge shi xiang-cai?* (Mandarin)

This which one-class. cop. coriander

‘Which one of these is coriander (flavour)?’

c. **ze** **dwe** **dʒi-lei** **si** **en-swi?** (Southern Min)

This which one-class. cop. coriander

‘Which one of these is coriander (flavour)?’

If we compare the monolingual sentences in (148b) and (148c), it is clear that there is a one-to-one correspondence between each word in these two sentences and that their word order is exactly the same. The morpheme order principle is then not applicable. Moreover, no outsider late system morpheme is found in (148a). This suggests that the system morpheme principle cannot be applied to (148a). Again, the two criteria in

the MLF model fail to identify the ML of my Mandarin/Southern Min data.

It might be argued that different forms of copula could be clues for an agreement relationship between the subject and verb, such as in English “a plural subject” always appears with “are” and “a singular subject” always occurs with “is”. However, this kind of subject-copula agreement relationship does not exist in Mandarin and Southern Min. Consider the Mandarin and Southern Min examples in (149).

(149) a. *ta shi wo tong-shi.* (Mandarin)

3s Cop. my colleague

‘He (she) is my colleague.’

b. *ta-men shi wo tong-shi.* (Mandarin)

3p. Cop. my colleagues.

‘They are my colleagues.’

c. *i ji wa don-fu.* (Southern Min)

3s cop. my colleague

‘He(she) is my colleague.’

d. *in ji wa don-fu.* (Southern Min)

3p cop. my colleagues

‘They are my colleagues.’

As the two Mandarin examples in (149a) and (149b) show, the copular verb “shi” can occur either with the subject *ta* ‘he/she’, a third person singular pronoun, or with *ta-men* ‘they’, the third person plural pronoun. This suggests that there is no agreement relationship between the subject and the copula in Mandarin. The same

analysis holds for the two Southern Min examples in (149c) and (149d). Both the singular subject *i* ‘he/she’ and the plural subject *in* ‘they’ are followed by the copula “*ji*”. This shows that in Southern Min also, the copula does not need to agree with the subject.

In summary, from the discussion above it is clear that the two important principles (i.e., the morpheme order principle and the system morpheme principle) of the MLF model cannot be used to identify the ML of the Mandarin/Southern Min bilingual data. The major reasons for this result may be that Mandarin and Southern Min share most of their syntactic structures, especially word order, and there is an absence of the outsider late system morphemes in both languages.

### *Empirical Evidence*

The theoretical problems of the MLF model were discussed above. The next step is to use empirical evidence to support my argument. The MLF model will be used to analyse the Mandarin/Southern Min data collected from three different groups: Group 3 – Educated below University Level/ Older People (aged 50 – 60); Group 4 – University Level/Older People (aged 50 – 60); Group 5 – University Students (aged 18 – 25). The results of analysis are shown in Table 5.

**Table 5: Results of the Application of the MLF Model to Mandarin/Southern Min Data**

	<b>Mandarin= ML</b>	<b>Southern Min=ML</b>	<b>Problematic data</b>	<b>Total Number of Bilingual CPS</b>
<b>Group 3</b>	?	3	97 (97%)	100
<b>Group 4</b>	?	6	94 (94%)	100
<b>Group 5</b>	?	8	92 (92%)	100

\* (The question mark ‘?’ represents the data whose ML could not be defined.)

As Table 5 shows, the MLF model was tested with three different sets of Mandarin/Southern Min data, 300 bilingual clauses in total. For Group 3, the MLF model was only applicable to three clauses (3%), and it failed to identify the MLs of the remaining ninety-seven bilingual clauses (97%). For Group 4, the MLF model was only applicable to six clauses (6%), and it could not identify the MLs of the other ninety-four clauses (94%). The results of Group 5 were similar. The MLs of only eight clauses could be identified, while those of the remaining ninety-two clauses (92%) could not be identified. Thus, over 90% of the bilingual clauses in each group had an unidentifiable ML.

Table 5 shows that there were some clauses whose ML could be identified in each group. All these clauses were A-Not-A questions suggested by Li and Thompson (1981). In section 4.8, I argued that the word order of the A-Not-A question is different in Mandarin and in Southern Min. The construction of the A-Not-A question in Mandarin is “verb + negation +verb”. However, in Southern Min, the construction is “*wu* ‘have’ + verb...+negation”. Now, consider the example in (150a), which is taken from one of the Mandarin/Southern Min data sets. Its monolingual versions in Mandarin and in Southern Min are provided in (150b) and (150c).

(150) a. **li** *wu* *ding* *zue* *hao-hua* **e** **bo**?

You have book most luxury nom neg

‘Have you booked the most luxurious (room)?’

(Mandarin/Southern; *Italic*= Mandarin; **Bold**=Southern Min)

b. *ni* *yo* *mei* *yo* *ding* *zue* *hao-hua* *de*? (Mandarin)

You have neg. have book most luxury nom.

‘Have you booked the most luxurious (room)?’

c. **li** **wu** **diŋ** **ʃon** **ho** **e** **bo**? (Southern Min)

You have book most luxury nom. neg.

‘Have you booked the most luxurious (room)?’

If we compare the three examples in (150), it is clear that the bilingual clause in (150a) has the same word order as the monolingual Southern Min clause in (150c). This is because (150a) shows the “*wu* ‘have’ + verb...+negation” construction of Southern Min but not the “verb + negation +verb” construction of Mandarin. Hence, according to the morpheme order principle, Southern Min is identified as the ML of (150a). In summary, although the MLF model is applicable to A-Not-A questions, it still does not apply to most of our Mandarin/Southern Min bilingual utterances (more than 90% in each of the three groups).

#### 6.2.2.2. Aspect Markers as Outsider Late System Morphemes

Myers-Scotton (personal communication) suggested that aspect markers (c.f. section 4.3) in Mandarin are outsider late system morphemes. Here, I shall use the Mandarin perfective aspect marker “le” and its equivalent expression, namely “liao” in Southern Min, as examples to examine whether the aspect markers in Mandarin could be treated as outsider late system morphemes. Consider my Mandarin examples in (151).

(151) a. *wo kan na ben shu.* (Mandarin)

I<sub>s</sub> read that Mclass2 book

‘I read that book.’

b. *wo kan le na ben shu.*

1s read asp.-perf. that Mclass2 book

‘I finish reading that book.’

c. *ta kan le na ben shu.*

3s read asp.-perf. that Mclass2 book

‘He/she finishes reading that book.’

d. *ta-men kan le na ben shu.*

3p read asp-perf that Mclass2 book

‘They finish reading that book.’

(151a) is a tenseless sentence, while the sentences in (151b) to (151d) all include the perfective aspect markers “le”. According to the system morpheme principle, the inflectional morpheme, which marks subject-verb agreement, is an outsider late system morpheme. If we examine (151b) to (151d), it is clear that regardless of whether the subject is singular, plural, male or female, the aspect marker “le” can co-occur with it without changes in form. This suggests that “le” occurs independently and there is no agreement relationship between the subject and “le”. Hence, I argue that the Mandarin aspect markers are not outsider late system morphemes. The analysis above holds for other Mandarin aspect markers, such as the durative aspect marker “zhe”, the experiential aspect marker “guo” and so on.

The equivalent particle of the Mandarin aspect marker “le” in Southern Min is “liao” (c.f. section 4.3.1.2.). Similar to the discussion of the Mandarin perfective aspect marker “le” above, there is no agreement relationship between the subject and the Southern Min perfective aspect marker “liao”. Hence, I also argue that the Southern Min perfective aspect marker “liao” is not an outsider late system morpheme. This is

illustrated in the Southern Min examples in (152) below.

(152) a. **wa kuā hi bun tse.** (Southern Min)

1s read that Sclass2 book

‘I read that book.’

b. **wa kuā liao hi bun tse.**

1s read asp-perf that Sclass2 book

‘I finish reading that book.’

c. **i kuā liao hi bun tse.**

3s read asp-perf that Sclass2 book

‘He/she finishes reading that book.’

d. **in kuā liao hi bun tse.**

3p read asp.-perf. that Sclass2 book

‘They finish reading that book.’

Moreover, in one set of the Mandarin/Southern Min data (i.e., Group 3), which includes one hundred bilingual CPs, only three (3%) were found to have aspect markers. The relative infrequency of aspect markers in our data means that they would not be a reliable indicator of the ML.

#### 6.2.2.3. Other Possible Criteria

In the discussion above, I proved that the MLF model failed to identify the ML of the Mandarin/Southern Min data. In this section, I shall use other criteria to see if they can solve the theoretical problems caused by the nature of Mandarin and Southern Min.



### *Early System Morphemes and Bridge Morphemes as New Criteria?*

The aim of this section is to seek new criteria based on the existing theory proposed by Myers-Scotton (2002a). Although the system morpheme principle states that only the outsider late system morpheme must come from the ML, based on the empirical evidence, Myers-Scotton (2002a, p.113) also states that “most of the early system morphemes and bridge morphemes do come from the ML”. According to another principle she proposes, namely the **uniform structure principle**<sup>19</sup>, Myers-Scotton (2002a, p.120) argues that “the uniform structure predicts early and bridge late system morphemes from the Matrix Language as the unmarked choice - just because it gives preference to keeping structure uniform across the CP”. For this reason, identifying the language source of the early system morphemes and bridge morphemes may be a way of identifying the ML.

### *Early System Morphemes*

To examine whether the language of the early system morpheme can be adopted as a criterion for identifying the ML, it is necessary to discuss the question: what are the early system morphemes in Mandarin and Southern Min? The definition of an early system morpheme is repeated as below:

“Early system morphemes depend on their heads for information about their forms...and are indirectly elected by their head content morphemes.”

(Myers-Scotton 2002a, p.75)

I would argue the classifiers in Mandarin and Southern Min are early system

---

<sup>19</sup> The Uniform Structure Principle: “A given constituent type in any language has a uniform abstract structure and the requirements of well-formedness for this constituent type must be observed whenever the constituent appears” (Myers-Scotton 2002a, p.120).

morphemes. As Li and Thompson (1981, p.105) argue, “the choice of classifier is determined by the noun”. In other words, to choose an appropriate classifier, one should consider the shape, size, animacy, or other relevant semantic features of the head noun being modified (c.f. Table 1 in section 4.4 which is repeated as Table 6. below).

**Table 6: Mandarin and Southern Min Classifiers with Glosses in English**

	Mandarin	Label	Southern Min	Classifier Number
<b>long and round</b>	zhi	Mclass 1	gi	Sclass 1
<b>Square/rectangular and thick</b>	ben	Mclass 2	bun	Sclass 2
<b>square and thin</b>	zhang	Mclass 3	diun	Sclass 3
<b>round body</b>	ke	Mclass 4	liab	Sclass 4
<b>human</b>	ming	Mclass5	mia	Sclass5
<b>animal</b>	tiao	Mclass6	jia	Sclass6
<b>building</b>	jian	Mclass7	gin	Sclass7
<b>vehicles on the land</b>	tai	Mclass8	dai	Sclass8
<b>long construction</b>	dao	Mclass9	də	Sclass9
<b>event</b>	hui	Mclass10	giam	Sclass10
<b>square, thin and small</b>	pian	Mclass11	de	Sclass11

To make this clearer, I repeat some examples from section 4.4, which are listed below:

(153) *yi-zhi bi* (shape-oriented) (*Mandarin*)

one-Mclass1 pen

‘one pen’

(154) *ɔ̃zi-bun tse.* (shape-oriented) (**Southern Min**)

one-Sclass2 book

‘one book.’

The Mandarin classifier “zhi” in example (153) is used because the noun, *bi* ‘pen’, which it modifies, is a long and round object. Similarly, the Southern Min classifier “bun” in (154) is used because *tse* ‘book’ is a rectangular and thick object. With reference to Myers-Scotton’s (2002a) definition, it can be assumed that the classifiers in (153) and (154) are indirectly selected by their head nouns. Thus, I argue that the classifiers in Mandarin and Southern Min are early system morphemes.

The next step is to examine whether the classifier can be treated as a new criterion to identify the ML. According to the uniform structure principle, the early and bridge system morphemes from the ML are preferred. In this case, the language that provides these two kinds of morphemes is likely to be the ML. Consider my Mandarin/Southern Min example in (155a). The monolingual Southern Min and Mandarin versions are provided in (155b) and (155c).

(155) a. **ɕʒɪ-de**      *jin-pian*    **zu**      **di** *pi-fu* *xia*.

one-Sclass11 microchip inject at skin below

‘One piece of microchip is put below the skin.’

(*Italic*=Mandarin; **Bold** = Southern Min)

b. *yi-pian*      *jin-pian*    *zhu-she* *dao* *pi-fu* *xia*. (Mandarin)

one-Mclass11 microchip inject at skin below

‘One piece of microchip is put below the skin.’

c. **ɕʒɪ-de**      **ɕʒian-pin**    **zu**      **di** **pue-hu** **e-bin**. (Southern Min)

one-Sclass11 microchip inject at skin below

‘One piece of microchip is put below the skin.’

The monolingual Mandarin example in (155b) and the monolingual Southern Min example in (155c) have exactly the same word order. Moreover, no outsider late system morphemes are found in these two examples. Thus, the two criteria of the MLF model, namely the morpheme order principle and the system morpheme principle, are unable to identify the ML of (155a). According to my new criterion, namely the uniform structure principle, the ML of (155a) would be Southern Min. This is because Southern Min provides the early system morpheme i.e., the classifier “de”. It is then argued that Southern Min provides the basic grammatical frame of (155a). The Mandarin nouns (i.e., *jin-pian* ‘microchip’ and *pi-fu* ‘skin’) and the post-position *xia* ‘below’ in Mandarin are EL elements.

More examples are provided in (156). (156a) is my Mandarin/Southern Min CS example, and its monolingual versions in Mandarin and in Southern Min are provided in (156b) and (156c). Again, the ML of the bilingual Mandarin/Southern Min example in (156a) is unable to be identified using the MLF model. This is because the monolingual Mandarin and Southern Min examples in (156b) and (156c) have the same word order, and no outsider system morphemes are found in the bilingual example in (156a). Since the classifier “liap” in (156a) is in Southern Min, we would identify Southern Min as the ML according to the uniform structure principle.

(156) a. **i**    *jing-zhui*    **sen**    **dʒɿ-liap**    *zhong-liao*.

his neck            grow one-Sclass4 tumour

‘There is a tumour on his neck.’

(*Italic*=Mandarin; **Bold** = Southern Min)

b. *ta jing-zhui zhang yi-ke zhong-liao.* (Mandarin)

his neck grow one-Mclass4 tumour

‘There is a tumour on his neck.’

c. **i am-guna sen ɕɿ-liap pai-mi.** (Southern Min)

his neck grow one-Sclass4 tumour

‘There is a tumour on his neck.’

At first sight, checking the early system morphemes (e.g., the classifiers) to identify the ML of a bilingual Mandarin/Southern Min seems to be a plausible solution. However, this criterion is not always available. In one set of data (i.e., Group 3), which includes one hundred bilingual Mandarin/Southern Min CPs, only fifteen were found to have classifiers. In other words, checking early system morphemes is not an ideal criterion because it cannot be applied to the great majority of my data (approximately 85%).

### *Bridge Morphemes*

It has been proved that checking the early system morpheme is not an adequate criterion to deal with all the data. The next step is to see whether the bridge morpheme can provide some useful clues for identifying the ML of a given bilingual CP. I shall begin my discussion by determining what the bridge morphemes are in Mandarin and in Southern Min. The definition of a bridge morpheme is repeated below.

“[A] Bridge morpheme connects content morphemes with each other without

reference to the properties of a head. e.g. possessive ‘of’ and ‘s.’”

(Myers-Scotton 2002a, p.75)

According to the definition above, it is clear that possessive markers are bridge morphemes. Hence, it can be assumed that the Mandarin possessive markers “de” and Southern Min “e” possessive markers are bridge morphemes. The examples in (157) are provided to illustrate how the possessive markers in Mandarin and Southern Min are used.

(157) a. *wo de xue-sheng hen nian-ching.* (Mandarin)

I poss. student very young

‘My students are very young.’

b. **wa e ha-ŋŋ ɕŋŋ ʃiao-len.** (Southern Min)

I poss. student very young

‘My students are very young.’

(*Italic*=Mandarin; **Bold** = Southern Min)

Following the uniform structure principle, we may hypothesize that if the bridge morpheme in a CP is from language A, it is likely that language A is the ML. However, in one set of my data (i.e., Group 3), only seven percent of bilingual CPs had bridge system morphemes. Therefore, the language of the bridge morpheme is not an adequate criterion to identify the ML either.

#### *Sentence Final Particles as a New Criterion?*

In the discussion in 3.7.3, it was mentioned that Kamwangamalu and Lee (1991)

adopted both lexical and structural criteria to test their Mandarin/English CS data. Although their observation was not based on the MLF model, the key notion of their approach is very similar, namely there is a ML in a given CS sentence. The structural criteria they used to identify the ML include: word order, verbless clauses, and sentence final particles. In the discussion above, it was showed that Mandarin and Southern Min have the same word order. The omission of the main verb is allowed both in Mandarin and in Southern Min. Thus, we can ignore their first two criteria (i.e. word order and verbless clauses), and merely examine whether checking the sentence final particles is useful to identify the ML of our Mandarin/Southern Min CS data. With reference to Li Kamwangamalu and Lee's (1991) argument, it could be hypothesised that if the sentence final particle of a given bilingual CP is in language A, then the ML is language A. This is illustrated by my Mandarin/Southern Min example in (158).

(158) **li** **mu** **ji** [ *be dai hui-qu e* **ni** ]<sub>CP</sub>?

you neg cop will take back nom part.-Ques.

'It is the thing you will take back, isn't it?'

(*Italic=Mandarin*; **Bold = Southern Min**)

Consider the Mandarin/Southern Min subordinate clause in the in the square brackets in (158). According to my hypothesis, the ML of this bilingual CP is Southern Min because the question particle "ni" comes from Southern Min. However, after testing this criterion with one set of our Mandarin/Southern Min data (i.e., Group 3), which includes one hundred bilingual CPs, two major problems were found. First, only thirty-three CPs (33%) were found to have sentence final particles. For those CPs

which did not have sentence final particles, the new criterion was not applicable. Second, as was discussed in section 4.7, some sentence final particles in Mandarin and in Southern Min have exactly the same forms and the same meanings. In this case, using this new criterion to identify the ML is also problematic. This is illustrated by my examples in (159).

(159)a. **gon** [**be** *dui-zhi da-xing le-se* **ga** *jia-ju* **la.**]

say will place big-size rubbish and furniture part.-impat.

‘(They) say (they) will place big rubbish and furniture (in the empty lands).’

(*Italic=Mandarin; Bold = Southern Min; **Bold, Italic Underlined** =*

unable to identify the language)

b. **goŋ** [**be** **kən** **dwa-gian** **bun-sə** **ga** **ga-gu** *la.*]

say will place big-size rubbish and furniture part.-impat.

‘(They) say (they) will place big rubbish and furniture (in the empty lands).’

(**Southern Min**)

c. *shuo* [*yao* *due-zhi da-xing le-se* *han* *jia-ju* *la.*]

Say will place big-size rubbish and furniture part.-impat.

‘(They) say (they) will place big rubbish and furniture (in the empty lands).’

(*Mandarin*)

As it can be seen in the monolingual examples in (159b) and (159c), the Southern Min particle “**la**” has exactly the same function and the same form as the Mandarin particle “*la*”. Thus, it is difficult to determine whether the particle “**la**” in the bilingual clause in (159a) belongs to Mandarin or Southern Min.



Also, consider the other example in (160a). (160b) and (160c) are its monolingual versions in Southern Min and Mandarin.

- (160)a. *yi*                    *nian-ling* **lue**        **sən**        o?  
 according to    age                down        count        part-Ques.  
 ‘Do (we) count (the amount of money he can earn) according to (his) age?’  
 (*Italic=Mandarin; Bold = Southern Min; Bold, Italic and Underlined =*  
 unable to identify the language)
- b. **dʒiao**                **ni-gi**        **lue**        **sən**        **o**? (Southern Min)  
 according to    age                down        count        part.-ques.  
 ‘Do (we) count (the amount of money he can earn) according to (his) age?’
- c. *yi*                    *nian-ling* *qu*        *suan*        *o*? (Mandarin)  
 according to    age                go            count        part.-ques.  
 ‘Do (we) count (the amount of money he can earn) according to (his) age?’

As (160b) and (160c) show, the Southern Min question particle “**o**” and the Mandarin question particle “*o*” have the same meaning and the same form. Hence, whether the question particle “o” in (160a) is Mandarin or Southern Min is unclear. To sum up, the criterion of checking sentence final particles still fails to identify the ML of most of the Mandarin/Southern Min data. This conclusion is drawn for two main reasons. First, not every bilingual CP has a sentence final particle. Second, many Southern Min and Mandarin sentence final particles have exactly the same form and the same function as each other. Thus, it is difficult to judge which language those particles belong to if they occur in a bilingual CP.

### 6.2.3 A Revised Version of the MLF Model

The results of the data analysis in section 6.2.1 suggested that the MLF model could be applied to unambiguously identify the MLs of all the Mandarin/Tsou bilingual clauses. Such a result supports Myers-Scotton's (2002a) argument that the two participating languages in codeswitching do not contribute equally. In other words, there is always one language, namely the ML, which contributes more in terms of setting up the basic morphosyntactic structure of a given bilingual clause.

Regarding the Mandarin/Southern Min data sets, the results of data analysis shown by Table 5 suggests that most of the MLs of the Mandarin/Southern Min bilingual clauses were unable to be identified. The major reason is that there is a lack of outsider system morphemes in both Mandarin and Southern Min, and the two languages share most of their syntactic structures. Hence, the two key criteria i.e., morpheme order principle and system morpheme principle were not applicable to the Mandarin/Southern CS data. However, stating that the MLF model is not applicable does not mean that my Mandarin/Southern CS data are violations or counter-examples to it. A real violation, if there is any, would be a Mandarin/Southern Min clause with Mandarin word order but Southern Min outsider system morphemes or vice versa. In fact, no such examples were found in the three Mandarin/Southern Min data sets, which include three hundred Mandarin/Southern Min bilingual clauses in total.

I argue that Myers-Scotton's (2002a) view that there is always a distinction between a matrix language and an embedded language in a given bilingual clause still holds for the Mandarin/Southern Min data. It is only because of the linguistic nature of Mandarin and Southern Min, which makes the identification of the ML more complex. According to the MLF model, the ML of a given bilingual clause is identified with

reference to the surface word order and the inflectional morphology of the participating languages. The language pairs to which the MLF model is successfully applied (i.e. the ML can be unambiguously defined) are typologically very different languages (e.g., Swahili/English, Myers-Scotton 1993/1997; German/English, Fuller 2000; Mandarin/English, Wei 2001; Spanish/English, Jake, Myers-Scotton, & Gross 2002; Welsh/English, Deuchar 2006) and Mandarin/Tsou as demonstrated in this thesis. Note that, one of the participating languages in each language pair is either an inflecting or an agglutinating language. If the two participating languages in a bilingual clause are typologically very similar languages (i.e., sharing most of their syntactic structure) and are both isolating languages (i.e., the number of inflectional morpheme is very limited), then it is more difficult to identify the language, which provides the morphosyntactic frame of the clause. Hence, when analysing such a language pair (e.g., Mandarin/Southern Min), I argue that criteria other than the morpheme order principle and the system morpheme principle are necessary for the identification of the ML. Hence, a revised version of the MLF model, as shown in Table 7, will be proposed.

**Table 7: A Revised Version of the MLF Model**

	<b>Criteria</b>		<b>if not applicable, go to</b>
1	Morpheme Order Principle	(To apply simultaneously)	2
	System Morpheme Principle		
2	Morpheme Counting Principle		3
3	Uniform Structure Principle		

This revised model is based on the original MLF model (i.e., the morpheme order principle and the system morpheme principle) with the incorporation of two new criteria (i.e., the morpheme counting principle and the uniform structure principle). In

the following discussion, I shall explain why the four criteria in the revised MLF model were adopted and how to apply them to analyse the Mandarin/Southern Min data.

#### 6.2.3.1. Stage 1: The Morpheme Order Principle and System Morpheme Principle

The first stage of the application of the revised MLF model proposed in the present study includes the two important constraints of the Myers-Scotton's (2002a) latest version of the MLF model, namely the morpheme order principle and the system morpheme principle. According to much current literature these two principles have been successfully applied to various language pairs. They also unambiguously identified the MLs of all the Mandarin/Tsou and some of the Mandarin/Southern Min bilingual clauses (i.e., the A-Not-A question). One of the research questions the present study intends to answer is whether there is any universally applicable CS model. The empirical success of these two principles has made the MLF model a very possible candidate for a CS model with universal applicability. Thus, I decide to reserve the two principles, namely the morpheme order principle and the system morpheme principle as the criteria, which need to be checked in the first stage of the application of the revised model shown in Table 7.

#### 6.2.3.2. Stage 2: The Morpheme Counting Principle

In the earlier sections of this chapter, I demonstrated that in its original form the MLF model is not applicable to the Mandarin/Southern Min data. Modifying this model so that early system morphemes and bridge morphemes are used as criteria to identify the ML seems to be plausible, but these two types of morphemes are not always available in my data. The other criterion, namely checking the sentence final particles to identify the ML, did not work, either. One criterion proposed to solve the problem

of the identification of the ML of a bilingual clause in which the two participating languages share most of their syntactic structure and have no outsider late system morphemes is the **morpheme counting principle**.

#### *Problems of the Morpheme Counting Principle*

The morpheme counting principle was used as the criterion for the identification of the ML in an early version of the MLF model proposed by Myers-Scotton (1993). The definition of the morpheme counting principle is shown below:

“The ML is the language of more morphemes in interaction types including intra-sentential CS.”

(Myers-Scotton 1993/1997, p. 68)

In other words, the morpheme counting principle is a frequency-based criterion. However, it is important to note that Myers-Scotton (1993, p.68) stated that “frequency counts must be based on a discourse sample; they offer no reliable evidence if they are performed on single sentences. Cultural borrowings from the EL for new objects and concepts are excluded from the count.” She also argued that sociolinguistic and psycholinguistic factors together needed to be taken into account when attempting to identify the ML. That is, the unmarked language choice (i.e., the most appropriate or most frequently used language in a given community) or the speaker’s dominant language (i.e., the language in which the speaker is most proficient) is usually identified as the ML by the morpheme frequency criterion. However, several problems with this earlier version of the MLF model were found either by other scholars or by Myers-Scotton herself.

The first problem raised by the morpheme counting principle is related to the typology of the languages involved. Muysken (2000, p.66) argues that “an agglutinating language like Swahili encodes many grammatical concepts (which are crucial structurally) with an overt morpheme, while isolating languages often do not”. Hence, simply counting the frequency of morphemes for the identification of the ML may have the problem of favouring the agglutinating languages. Second, the basic unit of analysis is not clear (Myers-Scotton 1997; 2002a). Myers-Scotton (1993) claimed that this principle must be applied to a discourse sample. However, she did not define what would constitute such a sample or how large such a sample should be. Third, Myers-Scotton (2002a) states that although the morpheme counting principle worked for her Swahili-English CS data, it does not hold for the CS data of other language pairs. The final problem is that to treat the ML as the same as the dominant language or an unmarked choice is problematic. ML is “a grammatically based construct” not something defined with reference to extra-linguistic factors (i.e., an individual’s proficiency in a language or personal choice of a certain language) (p.62). Because of the problems discussed above, the morpheme counting principle was abandoned by Myers-Scotton (2002a).

#### *Re-introduction of the Morpheme Counting Principle in Recent Research*

In a more recent application of the MLF model, Myers-Scotton’s (1993) earlier version of the MLF model was re-introduced by Paradis, Nicoladis, and Genesee (2000, p.258) for the attempt to investigate “whether bilingual children demonstrate a developmental shift in structural properties of CS from a non-adult to an adult-like system”. Their CS data produced by fifteen French/English bilingual children were collected from four different periods. The first period was from the age of one to two.

Then, their observations and recordings continued on a six-month interval basis, namely from the age of 2 to 2:6, 2:6 to 3 and 3 to 3:6.

Paradis et al. first adopted an earlier version of the MLF model proposed by Myer-Scotton (1993), namely using the morpheme counting principle and a discourse as the basic unit of analysis, to identify the MLs of their English/French CS data. In their research, a discourse referred to a stretch of utterances produced by the children in naturalistic play activities with their father alone and their mother alone, which were audio- and videotaped. In other words, if a parent usually communicates with his or her child in English in a discourse, then they would treat English as the expected ML on the grounds of frequency. Paradis et al. (2000) then adopted the two principles of a later version of the MLF model (Myers-Scotton 2000), namely the morpheme order principle and system morpheme principle, to identify the ML on a clause-by-clause basis, and examined whether English was also the ML and whether one or both principles are violated. The frequencies of violations of the morpheme order principle and system morpheme principle collected from different observation periods were then calculated and compared in order to observe the subjects' development of language proficiency in English and French. The authors assumed that if the frequency of violations of the morpheme order principle and the system morpheme principle decreased, the language proficiency of the subject had increased. This, they argued, could be interpreted as a sign of their subjects' growing awareness of the morphosyntactic constraints reflected in the two principles.

Two major arguments were put forward to explain their re-introduction of the earlier version of MLF model. First, since many two-morpheme utterances were produced by the bilingual children, a stretch of discourse rather than a bilingual CP was a more

appropriate unit of analysis of their CS data. Second, as all the subjects were under the age of four, their language proficiency in both English and French was still developing. Many violations of the morpheme order principle and system morpheme principle (e.g., English/French bilingual clauses with English word order but French outsider late system morphemes or vice versa) were found. In fact, those examples of violation were used to measure the children's language development in order to answer their research question. Thus, the identification of the ML would have been impossible if a bilingual CP had been the basic unit of analysis. For instance, a bilingual clause might have English word order but French outsider late system morphemes in a clear case of violation of the MLF model. In such a case, the identification of the ML would be difficult. If the ML were not able to be identified, then the language which is expected to supply the basic morphosyntactic framework would be unknown. Thus, it would be impossible to measure the development of the children's awareness of the morphosyntactic constraints.

To solve this problem Paradis et al. (2000, p.251) adopted Myers-Scotton's (1993, Chapter 3) argument that "the determination of the ML should be based on both sociolinguistic and psycholinguistic factor for a stretch of discourse". They reported that the one-parent, one-language strategy (i.e., each parent used one language when talking to their children) was adopted in all the participants' families. Thus, the language usually used in the interaction between one parent and his/her child, was the ML. Regarding the psycholinguistic factors, the ML was defined as "the language from which the majority of morphemes in the discourse sample are taken" (Paradis et al. 2000, p.251). They went on to state that since the parent's language is usually used as the major medium of communication and most of the morphemes produced by the child in each discourse was in the parent's language, the ML identified by either the



sociolinguistic or the psycholinguistic criterion was the language of the parent. Hence, counting the frequencies of the morphemes in a discourse could identify the ML clearly and disregard the ambiguity caused by individual violations of the MLF model.

#### *Re-introduction of the Morpheme Counting Principle into This Study*

In line with Paradis et al's (2000) study the morpheme counting principle is re-adopted in this study as one of the criteria to identify the ML, though for entirely different reasons. The first problem of this principle raised by Muysken (2000) is related to the issue of language typology, namely that the morpheme counting principle is likely to favour a language with a more complex morphological system as the ML. This is because, taking an agglutinating language as an example, the grammatical relations is overtly marked by the use of different morphemes, while such information in an isolating language is often realised in the context. However, Muysken's argument is not a problem with the Mandarin/Southern Min CS data. Both Mandarin and Southern Min are isolating languages, and therefore the problem of favouring a morphologically more complex language does not exist.

The other problem of the morpheme counting principle is that empirical evidence suggests that this principle does not always hold with the language pairs other than Swahili/English. However, I argue that such a result could probably be caused by the typology of the participating languages, namely the possibility of favouring the language with more complex morphological system proposed by Muysken (2000) discussed above. As Mandarin and Southern Min are both isolating languages, such a problem can therefore be minimized.

The final problem is that, according to the earlier version of the MLF model, the ML had to be identified with reference to both psycholinguistic and sociolinguistic factors. Myers-Scotton (2002a) argues that, although the dominant language in psycholinguistics or the unmarked language in sociolinguistics in many cases may also be the language which supplies the basic morphosyntactic frame of a given bilingual CP, namely the ML, they still differ. The ML is an abstract notion and “a grammatical construct”, while the dominant language or the unmarked choice is determined with reference to extra-linguistic factors. The present study acknowledges such a distinction and argues that the Mandarin/Southern Min CS data, like many other language pairs tested in the current literature, were produced through the abstract frame-building process set up by the ML. However, as has been stated repeatedly, it is the structural nature of Mandarin and Southern Min which make the role of the MLs of Mandarin/Southern Min CS data less overt from a grammatical point of view.

Hence, criteria other than the morpheme order principle and system morpheme principle are required. As Myers-Scotton (2002a, p.61) argues, although the morpheme counting principle does not always work, “the language that is the source of the grammatical frame often supplies more morphemes in a bilingual CP”, as well as in a stretch of bilingual discourse. In other words, the ML identified by the morpheme order principle and the system morpheme principle is usually the same as that identified by the morpheme counting principle. Thus, it is reasonable to use the morpheme counting principle to identify the ML when the other two principles do not apply. With reference to this argument, a frequency-based criterion (having a bilingual CP as the basic unit of analysis) to identify the ML seems to be a plausible solution to the problems caused by the linguistic natures of Mandarin and Southern Min.

I am aware that adopting the morpheme counting principle as one of the criteria to identify the ML may have a certain degree of arbitrariness. For instance, one may ask the questions: does the production of bilingual utterances really include a process of calculating the number of the morphemes? If the answer is yes, is this process of calculation a result of human being's innate language ability or a result of how people actually use the language? I admit that I am not ready to answer these questions. The adoption of the morpheme counting principle is largely on empirical grounds. That is, it can be used to identify the MLs of most of the Mandarin/Southern Min bilingual clauses, and therefore achieves significant empirical success. As discussed in Chapter 3, other prominent structural CS models, such as Poplack's (1980) free morpheme constraint and the equivalence constraint and Macswan's (2000; 2005) minimalist CS model, were abandoned by this study because of empirical problems. Since most of the morphemes in Mandarin and Southern Min are free morphemes, the free morpheme constraints and minimalist model (i.e. the PF Disjunction Theorem) then failed to be useful in analysing the Mandarin/Southern Min CS data. Neither was the original MLF model applicable to the Mandarin/Southern Min data. The problem, however, was solved by the morpheme counting principle.

In summary, the originally identified problems relating to the morpheme counting principle are found not to be problematic for the Mandarin/Southern Min data. On the contrary, this frequency-based criterion is likely to solve the problems of the identification of the ML caused by the linguistic natures of Mandarin and Southern Min, namely that both languages have no outsider late system morphemes and share most of their syntactic structure. Thus, the morpheme counting principle is re-adopted as an additional criterion to identify the ML in the revised MLF model proposed by

the present study.

#### 6.2.3.3. Stage 3: The Uniform Structure Principle

The final criterion in the revised MLF model proposed by the present study is the uniform structure principle whose definition is repeated below:

The Uniform Structure Principle: “A given constituent type in any language has a uniform abstract structure and the requirements of well-formedness for this constituent type must be observed whenever the constituent appears (Myers-Scotton 2002a, p.120)”.

As mentioned in section 6.2.2.3, the adoption of the uniform structure principle means that the early system morphemes (e.g., the classifiers) and bridge late system morphemes may be used in identifying the ML of a bilingual CP. As has been demonstrated, checking these two types of morphemes seems to be a plausible criterion to identify the ML, though they are not always available in the Mandarin/Southern Min data. Nevertheless, the uniform structure principle remains incorporated in the revised MLF model since it is useful when other criteria of the previous two stages of the application of the revised MLF model fail to identify the ML of a bilingual Mandarin/Southern Min CP.

#### 6.2.3.4. The Application of the Revised MLF Model

It is important to note that the four criteria in the revised MLF model should be applied in a specific order. The first two criteria (i.e., the morpheme order principle and the system morpheme principle) in stage 1 have to be applied simultaneously. If these two criteria fail to identify the ML of a given bilingual CP, then the third

criterion in stage 2 (i.e., the morpheme counting principle) should be used. If the ML is still not identified, the final criterion (i.e., the uniform structure principle) should be checked. The primary reason to apply these four criteria in a specific order is that the morpheme order principle and the system morpheme principle (i.e., the original MLF model) have been successfully applied to many different language pairs (including the Mandarin/Tsou data) in which the participating languages are typologically different (c.f. section 1.3.). Thus, it is important to give these two principles priority.

The third and fourth criteria are proposed specifically to analyse the language pairs such as Mandarin/Southern Min, both are isolating languages and share many similarities in terms of word order and the absence of the outsider late system morphemes. The reason why the morpheme counting principle should be applied before the uniform structure principle is that the early system morphemes and bridge morphemes which are not always available (c.f. section 6.2.2.3.) in the Mandarin/Southern Min corpus. Therefore, it is not reasonable to treat the uniform structure principle as equal to the morpheme counting principle. Moreover, it is sometimes found that only the classifier phrase is switched in a bilingual clause. In such a case, it is not reasonable to claim that the ML is the language of the classifier, and the EL is the language, which provides most of the morphemes. Consider the underlined bilingual clause in the square brackets in (161).

- (161) **nən-e**     **lon-di**            **ga-tse, yo**    *bo-shi*    *shuo* [ ni    yo  
 two-class. both-asp,dur    teach    and    neg.-be    say    you    have  
hen    da    de    **hi-lə.**] CP  
 very    big    nom. that-class.

'Both of them are teachers, and they don't have a very big (business).'

(*Italic*= Mandarin; **Bold**=Southern Min)

In the underlined subordinate clause in (161), only the classifier phrase “hi-lə” is in Southern Min and the rest of the morphemes are all in Mandarin. It is not reasonable to claim that Southern Min is the ML of this bilingual clause. Thus, it is better to put the morpheme counting principle in a higher hierarchical position than the uniform structure principle. To illustrate how the revised MLF model is applied to identify the ML of a given bilingual clause, consider my Mandarin/Southern Min example in (162a). Its monolingual versions in Southern Min and Mandarin are provided in (162b) and (162c).

(162) a. *i er-zi* **a-bue bi-iap** e **kuan o.**

her son not yet graduate nom seem part-affirm

'Her son has not graduated yet.'

(*Italic*=Mandarin; **Bold** = Southern Min)

b. **i giã a-bue bi-iap** e **kuan o.** (Southern Min)

her son not yet graduate nom shape part-affirm

'Her son has not graduated yet.'

c. *ta er-zi hai-mei bi-ye de yang-zi o.* (Mandarin)

her son not yet graduate nom shape part.-affirm

'Her son has not graduated yet.'

To identify the ML of the bilingual clause in (162), the first two criteria in Table 7 (i.e., the morpheme order principle and the system morpheme principle) needs to be

checked. Since the monolingual clauses in (162b) and (162c) have the same word order and no outsider late system morphemes are found in (162a), neither the morpheme order principle nor the system morpheme principle is applicable. In this case, the third criterion in Table 7 (i.e., the morpheme order principle) is used to identify the ML of (162a). In (162a), only the word *er-zi* ‘son’ is from Mandarin, while all other words are in Southern Min. According to the morpheme counting principle, the language that supplies more morphemes is the ML. Thus, Southern Min is identified as the ML of the bilingual clause in (162a).

Sometimes it is found that, in a bilingual Mandarin/Southern Min clause, the two participating languages provide an almost equal number of morphemes. This means that the ML is unable to be identified with reference to the morpheme order principle, the system morpheme principle, and the morpheme counting principle in Table 7. In this case, the final criterion, namely the uniform structure principle will be used. According to this principle, the early and bridge system morphemes (i.e., the classifiers and the possessive markers in Mandarin and Southern Min) from the ML are preferred. In section 6.2.2.3, I argued that the classifiers in Mandarin and Southern Min are early system morphemes and possessive markers are bridge morphemes. Hence, it is assumed that the language, which provides these two types of morpheme, is the ML. To illustrate how the fourth criterion, the uniform structure principle, in the revised MLF model is applied, consider the example in (156), repeated here as (163).

(163)**i**    *jing-zhui*    **sen**    **ɕʒi-liab**    *zhong-liao*. (Mandarin/Southern Min)  
 His neck            grow    one-Sclass4    tumour  
 ‘There is one tumour on his neck.’

(*Italic*= Mandarin; **Bold**=Southern Min)

In (163), the number of words provided by both languages is nearly equal (i.e., two in Mandarin, and three in Southern Min). It would be problematic if Mandarin is identified as the ML simply because it provides one word more than does Southern Min. In this case, the morpheme counting principle is not plausible<sup>20</sup>, and the fourth criterion, the uniform structure principle, should be checked. In (163), the language of the early system morpheme, namely the classifier “**liab**” is provided by Southern Min. Thus, Southern Min is identified as the ML. A more complex example is given in (164).

(164) **a**    **wa**    **mu**    **zai**        [ yi e        bode    shi    she-me. ]

DM I        neg. know        he poss.    purpose cop. what

‘Well...I don’t know what his purpose is.’

(*Italic= Mandarin; Bold=Southern Min*)

(164) is a complex clause. Southern Min is identified as the ML of the entire complex clause, for it provides most of the words. However, the morpheme counting principle is not applicable to the underlined subordinate clause because the number of words provided by Mandarin and Southern Min is almost equal. Thus, the fourth criterion must be used. Since the possessive marker *e* (i.e., a bridge morpheme) is provided by Southern Min, then Southern Min is identified as the ML of this subordinate clause.

The revised MLF model shown by Table 7 was applied to three sets of Mandarin/Southern Min CS data collected from Groups 3, 4, and 5 (three hundred

<sup>20</sup> In this thesis, I decided arbitrarily that the third criterion (the morpheme counting principle) is only applicable when one participating language provides at least two more words than the other language.



bilingual clauses in total). It should be noted that a complex clause, such as (160), was analysed twice because it contains two bilingual CPs, namely the entire complex clause and the subordinate clause. The overall results of analysis are shown by Table 8.

**Table 8: The Results of Applying the Revised MLF Model to Mandarin/Southern Min Data**

Matrix Language	Group 3			Group 4			Group 5		
	Mandarin	Southern Min	Total	Mandarin	Southern Min	Total	Mandarin	Southern Min	Total
Number of Bilingual Clauses	17	79	96	27	69	96	24	74	98
Number of Clauses whose MLs cannot be identified	4		4	4		4	2		2
Total			100			100			100

As Table 8 shows, the ML of over ninety percent of the bilingual Mandarin/Southern Min clauses in each group was successfully identified by the revised MLF model. However, some problematic data to which the revised MLF model was not applicable were found. (165a) is one example of the problematic data. Its monolingual Mandarin and Southern Min versions are provided in (165b) and (165c).

(165)a. *yi*                    *nian-ling* **lwe**    **sən**    o? (Mandarin/Southern Mine)

according to    age                    down    count    part.-ques

‘Do we count the amount of money he can get according to his age?’

(*Italic*=Mandarin; **Bold** = Southern Min; **Bold, Italic Underlined** =

unable to identify the language)

b. *yi*                    *nian-ling* *qu*                    *suan*    *o?* (Mandarin)

according to    age                    go                    count    part.-ques.

‘Do we count the amount of money he can get according to his age?’

c. **dʒiə**            **ni-gi**            **lwe**            **sən**            **o?** (Southern Min)

according to    age            down            count    part.-ques.

‘Do we count the amount of money he can get according to his age?’

To identify the ML of the bilingual clause in (165a), the first two criteria of the revised MLF model shown by Table 7, namely the morpheme order principle (i.e., to check word order) and system morpheme principle (i.e., to check the outsider late system morphemes, such as subject-verb agreement affixes or case markers) have to be checked. If we compare the bilingual clause in (165a) to the monolingual clauses in (165b) and (165c), it is clear that no difference is found in terms of their word order. In addition to that, no late outsider system morpheme is found in (165a). Hence, the first criterion of the MLF model is not applicable to (165a).

Then, we move to the third criterion, namely the morpheme counting principle, which argues that the language that supplies the more morphemes is the ML. However, in (164a) two morphemes (i.e., *according to* ‘yi’ and *age* ‘nian-ling’) are in Mandarin and two morphemes (i.e., *down* ‘lwe’ and *count* ‘sən’) are in Southern Min. The language of the bold, italic and underlined sentence final question particle “o” cannot be identified because, as the monolingual examples in (165b) and (165c) show, it could either be Mandarin or Southern Min. Since Mandarin and Southern Min provide the same number of morphemes in (165a), the morpheme counting principle is therefore not applicable either.

Finally, the fourth criterion, namely the uniform structure principle, should be

checked to identify the ML of (165a). As mentioned above, the classifiers in Mandarin and Southern Min are identified as early system morphemes, and the possessive marker *de* is identified as a bridge morpheme. In (165a), neither of these two kinds of morphemes is found, and this indicates that the uniform structure principle is not applicable, either.

To sum up, the four criteria of the revised MLF model shown by Table 7 were found not to be applicable to the problematic data. As Table 8 shows, only very small proportions of each of the three sets of Mandarin/Southern Min CS data were found to be problematic (Group 3= 4%; Group 4 = 4%; Group 5= 2%). This shows that the revised model works successfully in terms of identifying the ML of the great majority of the Mandarin/Southern Min data.

### 6.3. Testing the Typology Approach

In the previous section, I demonstrated some theoretical and empirical problems of the original MLF model. Although I proposed a revised version of the MLF model, it can only be applied to examine intra-clausal CS. Muysken's (2000) typological approach, however, provides a wider approach to CS. The typological approach covers intra-clausal CS, inter-clausal CS, and switches that occur between different turns. Hence, it may be interesting to test his model and see how it works when applied to the Mandarin/Tsou and Mandarin/Southern Min data.

#### 6.3.1. Review of the Typology Approach

The aim of this section is to test Muysken's (2000) typological approach. Before applying his model to my data, it is necessary to have an overview of the major claims of his theory. Muysken (2000) proposes three different CS patterns, namely, insertion,

alternation, and congruent lexicalization (c.f. section 3.4). He argues that insertion refers to lexical items or constituents from one language that are inserted into the structure of the other language (i.e., the matrix language). This is illustrated by Myers-Scotton's (1993/1997, p.81) Swahili/English CS example in (166). The English words *sure*, *suspect* and *week* are inserted in to a grammatical structure provided by Swahili.

(166) Na kweli, hata mini si-ko *sure* lakini n-a-*suspect* i-ta-kuwa *week* kesho.

‘Well, even I am not sure, but I suspect it will be next week.’

(Swahili/English; Myers-Scotton 1993/1997, p.81)

Muysken (2000) argues that alternation refers to the type of CS in which the grammatical structures of the two participating languages remain separate. Consider Treffers-Daller's (1994, p.213) French/Dutch CS data in (167), which, according to Muysken, is a typical case of alternation. In (167), the grammatical structures of French and Dutch occur simultaneously and separately within one sentence.

(167) Je téléphoné à Chantal, he, *meestal voor commieskes te*

I call to Chantal, hm mostly for shopping to  
*doen en eten.*

do and food

‘I call Chantal to go shopping and get food.’

(French/Dutch; Treffers-Daller 1994, p.213)

To recall the definition given in section 3.4.3, congruent lexicalization refers to the CS pattern in which different languages/language varieties are involved, and these

languages may, entirely or partially, share their grammatical structure (Muysken 2000). Muysken cites Giesbers' (1989, p.249) Ottersum dialect /standard Dutch CS data, as provided in (168), for illustration. Muysken (2000) argues that despite some phonological and lexical differences, the major syntactic structures of these two varieties of Dutch have almost no difference. Hence, the speakers can switch back and forth between lexicons in these two languages.

(168)...nee *onder leiding van jou* gedaan, da gij daor de keuke *kent*, op de  
Roepaan...

‘... no/under your direction/done, that you know the kitchen there, on the  
Roepaan...’

(*Italic and bold* = standard Dutch; Normal print = Ottersum dialect)

Muysken also argues that these three CS patterns can be linked to different linguistic and extra-linguistic factors. His arguments are illustrated by Table 9 (Deuchar 2004; summarised from Muysken 2000: 8-9).

**Table 9: Muysken's View of the Relationship between Code-switching Patterns and Extra-linguistic Factors**

<b>Code-switching pattern</b>	<b>Linguistic factors favouring this pattern</b>	<b>Extra-linguistic factors favouring this pattern</b>
<b>Insertion</b>	Typological distance	Colonial settings Recent migrant communities Asymmetry in speaker's proficiency in two languages
<b>Alternation</b>	Typological distance	Stable bilingual communities Tradition of language separation
<b>Congruent lexicalization</b>	Typologically similar languages	Two languages have roughly equal prestige No tradition of overt language separation Dialect/standard and post-creole continua

Table 9 shows the relationship between the CS patterns and linguistic/extra-linguistic factors. From a linguistic perspective, insertion and alternation are likely to occur when the participating languages in a given CS utterance are typologically different, while congruent lexicalization usually occurs when the languages involved are typologically similar. From an extra-linguistic point of view, insertion often occurs in colonial settings or in recent migrant communities. Alternation is likely to occur in a stable bilingual community and congruent lexicalization is commoner in a dialect/standard and post-creole continua.

### 6.3.2. Theoretical Problem - How to Define A Switch?

Muysken's (2000) typology covers a wider range of perspectives on CS if compared to the MLF model. It not only looks at intra-clausal CS, but also examines inter-clausal CS and CS that occurs between different turns. However, in Deuchar,



'cooperative association' will be a switch because (169) begins with the Southern Min pronoun *yi* 'he' and the discourse marker *hi-le* 'well'. The Southern Min possessive marker "e" should also be treated as a switch because the word before it is in Mandarin. However, the sequential approach is problematic in two ways. First, Muysken's (2000) model also deals with the switches that occur between different turns. In this case, if the last element in the previous turn is in Mandarin, then the Southern Min elements *yi* 'he' and *hi-le* 'well' in (169) should be treated as switches as well. Since Muysken's model does not have a clear unit of analysis, how a switch is to be defined is unclear. Second, every time the language of an element is different, it may be counted as a switch. The result of such an approach would be that almost every single element in this clause should be treated as a switch.

To sum up, Muysken's (2000) approach not only covers intra-clausal switches, but also inter-clausal switches and even switches between different turns. For this reason, determining a switch is more complex, and the basic unit of analysis becomes even more elusive. Since the basic unit of analysis of Muysken's (2000) model is not clear, it is difficult to define what a switch is. If a switch is hard to identify, the diagnostic features (c.f. section 3.4.) designed to identify the CS pattern of a given switch will be impossible to apply. This is, I argue, is one of the major problems of the typological model.

### 6.3.3. A Revised Typology Approach

Having demonstrated some theoretical problems with Muysken's (2000) typology approach, in the following discussion, I will attempt to revise his original model with reference to the Mandarin/Tsou and Mandarin/Southern Min data.



### 6.3.3.1. CP as A Basic Unit of Analysis

Since the difficulty of identifying a basic unit of analysis is the major problem in Muysken's (2000) typology approach, I decided here, in line with Deuchar, Muysken, and Wang (forthcoming), to analyse our data on a clause basis (CP). This implies that some important notions of the MLF model will be adopted, especially assuming that there is a ML in a CP. If we can identify of the ML of a CP, then the problem of how to identify a switch can be solved. For example, if language A is the ML of a bilingual CP, then any other elements in language B can be identified as switches. However, this raises another question: how is the ML of a CP defined? This is not a problem for the Mandarin/Tsou data (c.f. 6.2.1). However, in section 6.2.2, I demonstrated that Mandarin and Southern Min have the same SVO order, and there is an absence of outsider late system morphemes in both languages. Hence, the two principles (i.e., the morpheme order principle and the system morpheme principle) of the original MLF model are not applicable to the Mandarin/Southern Min data. For this reason, the revised MLF model shown by Table 7, repeated here as Table 10, will be adopted.

**Table 10: A Revised Version of the MLF Model**

	<b>Criteria</b>		<b>if not applicable, go to</b>
1	Morpheme Order Principle	(To apply simultaneously)	2
	System Morpheme Principle		
2	Morpheme Counting Principle		3
3	Uniform Structure Principle		

The method of identifying intra-clausal switches was demonstrated in section 6.2.3.4. Hence, the following discussion will focus on how the inter-clausal switches may be identified.

### *Inter-clausal Switches*

When dealing with inter-clausal switches, the first step is to identify the MLs of the two clauses. If the ML of the second clause is different from that of the first clause, then we count the second clause as a switch. This sequential approach is in line with Deuchar, Muysken and Wang (forthcoming). Consider the Mandarin/Southern Min example in (170).

(170) [ *tai-bi*                    *yi-ge*      *yi-qian*            *liang-bai*      *kuai.* ]<sup>1</sup>    [ **wa**  
Taiwanese-dollar one-class one-thousand two-hundred dollar I  
**kwai hwun-ki.** ]<sup>2</sup>  
almost faint

‘One (water melon) is one thousand and two hundred dollar. Oh my god!’

(*Italic*= Mandarin; **Bold**=Southern Min)

Clause 1 and clause 2 in (170) are both monolingual, so to identify their MLs is straightforward. The ML of the first clause is Mandarin, and that of the second clause is Southern Min. Since there is a change of the ML in the second clause, the entirety of clause 2 is then identified as an inter-clausal switch. In my corpus, it is sometimes found that the two clauses to be analysed are both bilingual, as shown in (171). In such a case, identifying the MLs and the switches is more complex.

(171) [ *zhe-li you xi-gua*            **ne** ]<sup>1</sup>            [ *xi-gua*            **dʒin-ma zo**  
here have watermelon part-excl watermelon now very  
**gui ne.** ]<sup>2</sup>  
expensive part.-excl

‘Here are some watermelons! Watermelon is very expensive now!’

(*Italic*= Mandarin; **Bold**=Southern Min)

In (171), Mandarin is identified as the ML of clause 1, for it provides most of the elements in this clause. The only Southern Min element, namely the sentence final particle “**ne**” in this clause is then treated as an intra-clausal switch. In clause 2, Southern Min provides the majority of elements. Therefore, Southern Min is identified as the ML. The only Mandarin element *xi-gua* ‘watermelon’ is then identified as an intra-clausal switch. If we compare clause 1 with clause 2, it is clear that there is a change of the ML. Thus, the entire clause 2 is treated as an inter-clausal switch.

#### 6.3.3.2. The Application of Muysken’s Features

By the application of our clause-based method of analysis, we could clearly identify all the intra-clausal and inter-clausal switches in our two sets of Mandarin/Tsou data and three sets of Mandarin/Southern Min data. The next step is to analyse each switch by using the different diagnostic features proposed by Muysken (2000) (c.f. section 3.4).

#### *A Quantitative Analysis*

Table 11 is cited from Muysken (2000, p.230), and includes all the diagnostic features that may be used to analyse the CS patterns.

**Table 11: Diagnostic Features of the Three Patterns of CS**

	<b>Insertion</b>	<b>Alternation</b>	<b>Congruent lexicalization</b>
<b>constituency</b>			
single constituent	+	0	0

several constituents	-	+	0
non-constituent	-	-	+
nested a b a	+	-	0
not nested a b a	-	+	+

### **element switched**

diverse switches	-	0	+
long constituent	-	+	-
complex constituent	-	+	-
content word	+	-	-
function word	-	-	+
adverb, conjunction	-	+	-
selected element	+	-	+
emblematic or tag	-	+	0

### **switch site**

major clause boundary	0	+	0
peripheral	0	+	0
embedding in discourse	0	+	0
flagging	-	+	-
dummy word insertion	+	0	-
bidirectional switching	-	+	+

### **properties**

linear equivalence	0	+	+
telegraphic mixing	+	-	-
morphol. integration	+	-	+
doubling	-	+	-
homophonous diamorphs	0	-	+
triggering	0	0	+
mixed collocations	0	-	+
self-corrections	-	+	-

---

“+” = indicative of a specific pattern

“-” = counter-indicative of a specific pattern

“0” = indicates that the feature is neutral to that pattern.

As Table 11 shows, the diagnostic features proposed by Muysken can be divided into four different categories, namely constituency, element switched, switch site, and properties. The “+” in the table indicates that this feature favours a certain CS pattern, while “-” represents a certain pattern which this feature does not favour. “0” indicates that the feature is neutral to that pattern.

Based on Table 11, Deuchar (2004) developed a scoring spreadsheet, which can identify the CS pattern of a given switch by calculating the pattern-specific scores. To illustrate her scoring procedures, consider the Mandarin/Southern Min example in (172).

(172) **an-na** *gu-li*        **in**        **e.**  
      how    encourage    them    nom  
      ‘How to encourage them?’

(*Italic*= Mandarin; **Bold**=Southern Min)

First, the ML and the switched elements should be identified. Southern Min is identified as the ML of (172), as it provides the majority of the morphemes. The Mandarin word *gu-li* ‘encourage’ is therefore identified as a switch. Second, each diagnostic feature listed in Table 11 will be applied to analyse this Mandarin switch. The detailed procedures of applying the features to this switch are illustrated by Table 12.

Table 12: The Quantitative Analysis of Data Using Muysken's Features

Row 1	Column 1	Column 2	Column 3	Column 4	Col. 5	Col. 6	Col. 7	Col. 8
Row 2	MUYSKEN'S Features	INSERTION	ALTERNATION	CONGRUENT LEXICALIZATION	example: gu-li	<u>Score</u> <u>on</u> <u>INS</u>	<u>Score</u> <u>on</u> <u>ALT</u>	<u>Score</u> <u>on</u> <u>CON</u>
Row 3	single constituent	+	0	0	+	1	0	0
Row 4	several constituents	-	+	0	-	1	-1	0
Row 6	non-constituent	-	-	+	-	1	1	-1
Row 7	nested aba	+	-	0	+	1	-1	0
Row 8	non-nested aba	-	+	+	-	1	-1	-1
Row 9	DIVERSE SWITCHES	-	0	+	Type of switch: Past participle			
Row 10	long constituent	-	+	-	-	1	-1	1
Row 11	complex constituent	-	+	-	-	1	-1	1
Row 12	content word	+	-	-	+	1	-1	-1
Row 13	function word	-	-	+	-	1	1	-1
Row 14	adverb, conjunction	-	+	-	-	1	-1	1
Row 15	selected element	+	-	+	-	-1	1	-1
Row 16	emblematic or tag	-	+	0	-	1	-1	0

Row 17	major clause boundary	0	+	0	-	0	-1	0
Row 18	peripheral	0	+	0	-	0	-1	0
Row 19	embedding in discourse	0	+	0	0	0	0	0
Row 20	flagging	-	+	-	-	1	-1	1
Row 21	dummy word insertion	+	0	-	-	-1	0	1
Row 22	BI-DIRECTIONAL SWITCHING	-	+	+	Switch into Mandarin			
Row 23	linear equivalence	0	+	+	+	0	1	1
Row 24	telegraphic mixing	+	-	-	-	-1	1	1
Row 25	morpho-gical integration	+	-	+	0	0	0	0
Row 26	doubling	-	+	-	-	1	-1	1
Row 27	homophonous diamorphs	0	0	0	0	0	0	0
Row 28	triggering	0	0	+	0	0	0	0
Row 29	mixed collocations	0	-	+	-	0	1	-1
Row 30	self-corrections	-	+	-	-	1	-1	1
Row 31					Total score	11	-7	3

In Table 12, the cell at column 5, row 2 is the Mandarin switch *gu-li* ‘encourage’. Since this verb matches the feature *single constituent* in column 1 & row 3, a positive value “+” is given, as shown in the cell in column 5 & row 3. Then, we compare this cell to other cells in column 2 & row 3, column 3 & row 3, and column 4 & row 5. These cells represent the values, which favour or do not favour the three CS patterns (i.e., insertion, alternation, and congruent lexicalization). Since the value in the cell in column 5 & row 3 matches the positive value in the cell in column 2 & row 3, a score of 1 is given, as shown in the cell in column 6 & row 3. The values in the cells in column 3 & row 3 and column 4 & row 3 are “0”, a score of 0 is given in the cells in column 7 & row 3 and column 8 & row 3.

If we move to the second feature *several constituents* in column 1 & row 4, it is clear that the verb “*gu-li*” does not match this feature. Therefore, a negative value “-” is given, as shown in the cell in column 5 & row 4. Then, we compare this negative value to those in the cells in column 2 & row 4, column 3 & row 4, and column 4 & row 4. Since the cell in column 5 & row 4 matches the negative value of the cell in column 2 & row 4, a score of 1 is given, as shown in cell in column 6 & row 4. However, it does not match the positive value of the cell in column 3 & row 4, and therefore we give it a score of -1, as shown in the cell in column 7 & row 4. Finally, the value in column 4 & row 4 is ‘0’. Thus, a score of 0 is given, as shown in the cell in column 8 & row 4.

It is important to note that if a feature is not applicable to the identified switch, a score of 0 is given. For instance, the feature *morphological integration* in column 1 & row 25 is not applicable to the Mandarin switch *gu-li* ‘encourage’. It is, then, given a score of 0, as shown in the cell in column 5 & row 25, column 6 & row 25, column 6 & row



26 and column 6 & row 26. In fact *morphological integration* is not applicable to almost all of the Mandarin/Southern Min data. This is because both Mandarin and Southern Min are both isolating languages in which most of their morphemes can occur independently. Thus, *morphological integration* is very rare in my Mandarin/Southern Min corpus.

By following the same procedure, all the features were applied to this Mandarin switch *gu-li*. The overall score is shown at the bottom of the table. Insertion receives a total score of 11. Alternation receives a score of -7. Congruent Lexicalization receives a score of 3. It is clear that insertion gets the highest score, and therefore it is concluded that the CS pattern of this Mandarin switch *gu-li* 'encourage' is insertion.

### *Results*

Adopting the methodology described above, five different sets of data were analysed including two sets of Mandarin/Tsou data, and three sets of Mandarin/Southern Min data. These five groups were differentiated by the subjects' age or age and educational level combined. The characteristics of these five groups were described in section 5.2, and are repeated below.

#### Mandarin/Tsou Groups:

Group 1: Old Tsou People — 55 ~ 70 years old

Group 2: Young Tsou People — 25~ 40 years old

#### Mandarin/Southern Min Groups:

Group 3: Under University Level/Elders — 50 ~ 60 years old

Group 4: University Level/Elders — 50 ~ 60 years old

Group 5: University Students — 18 ~ 25 years old

One hundred switches were available from Groups 1, 3, 4, and 5. In Group 2 Mandarin seems to replace Tsou as the major language for communication, since in approximately six-hours of recording only thirty switches were collected. Hence, four hundred and thirty switches in total were analysed using Deuchar's (2004) spreadsheet. The results are shown in Table 13.

**Table 13: The Overall Results of Mandarin/Tsou and Mandarin/Southern Min Data**

<b>Scores on Switching Patterns</b>	<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>	<b>Group 4</b>	<b>Group 5</b>
<b>Insertion</b>	543	208	494	500	631
<b>Alternation</b>	-505	-149	-440	-424	-498
<b>Congruent Lexicalization</b>	517	168	638	568	557
<b>Dominant Pattern</b>	<b>I</b>	<b>I</b>	<b>C</b>	<b>C</b>	<b>I</b>

(I= Insertion; C=Congruent Lexicalization)

The results of the Mandarin/Tsou data collected from Group 1 will be examined firstly. Table 13 suggests that the dominant pattern of Group 1 is insertion, for it receives the highest score of 543 if compared with the scores received by the other two CS patterns. The secondary pattern is congruent lexicalization which receives a score of 517, and alternation is the lowest with a score of just -505. For Group 2, the dominant pattern is also insertion with a score of 208. Congruent lexicalization receives the second highest score 168. Alternation is still the lowest and only achieves a score of -149. To sum up, the dominant CS pattern of the two sets of Mandarin/Tsou data is insertion.

For the Mandarin/Southern Min data (i.e., Groups 3, 4, & 5), the dominant pattern in Group 3 is congruent lexicalization, which receives the highest score of 638. Insertion comes the second and receives a score of 494. Alternation is the lowest and merely scores -440. The dominant pattern of Group 4 is also congruent lexicalization (with a score of 568). The score of the secondary pattern, namely insertion, is 500, which is very close to that of the dominant pattern. Alternation comes last and gets a score of -424. For Group 5, in contrast to Groups 3 and 4, insertion is the dominant pattern and receives a score of 631. The secondary pattern of this group is congruent lexicalization, with a score of 557. Alternation is still the lowest and only receives a score of -498.

#### *Theoretical Problems of Muysken's Model*

At first sight, the overall results in Table 13 look promising, for there is always a dominant CS pattern for each group. However, some theoretical problems, from both linguistic and extra-linguistic points of views, are found if we compare the results in Table 13 to the predications of Muysken's (2000) model in Table 9, repeated as Table 14 below.

**Table 14: Muysken's View of The Relationship between Code-switching Patterns and Extra-linguistic Factors**

<b>Code-switching pattern</b>	<b>Linguistic factors favouring this pattern</b>	<b>Extra-linguistic factors favouring this pattern</b>
<b>Insertion</b>	Typological distance	Colonial settings Recent migrant communities Asymmetry in speaker's proficiency in two languages
<b>Alternation</b>	Typological distance	Stable bilingual communities Tradition of language separation
<b>Congruent lexicalization</b>	Typologically similar languages	Two languages have roughly equal prestige No tradition of overt language separation Dialect/standard and post-creole continua

According to Table 14, insertion, from a linguistic perspective, often occurs when the two participating languages are typologically different, while congruent lexicalization is likely to occur when the two languages are typologically similar. Mandarin (SVO) and Tsou (VOS) are typologically different languages, while Mandarin and Southern Min (SVO) are typologically similar languages. Hence, we would expect insertion as the major pattern in the Mandarin/Tsou data, and congruent lexicalization as the dominant pattern in the Mandarin/Southern Min data. However, as the results in Table 13 show, the dominant pattern of the Mandarin/Southern Min data in Group 5 is insertion and not congruent lexicalization, which is contradictory to the prediction in Table 14. Furthermore, the scores of insertion and congruent lexicalization in some groups (i.e. Groups 1, 2, and 4), as Table 13 shows, are very similar. For instance,

insertion receives a score of 543 and congruent lexicalization gets a score of 517 in Group 1. For Group 2, the score for insertion is 208 and that of congruent lexicalization is 168. For Group 4, the score received by insertion is 500 and the score received by congruent lexicalization is 568. In other words, besides the dominant pattern, there is always a secondary pattern and the differences in scores they receive are not always significant. Hence, it is questionable to argue that each CS favours a certain linguistic factor.

From an extra-linguistic perspective, Muysken's (2000) model also has some problems. According to Table 14, insertion often occurs in a colonial setting and congruent lexicalization is likely to occur when the two participating languages are a dialect and its standard language. An endangered language like Tsou may be considered as existing in a colonial setting. The relationship between Southern Min and Mandarin can certainly be described as a dialect to a standard language. Therefore, we would predict that the dominant pattern of Mandarin/Tsou data would be insertion and that of Mandarin/Southern Min would be congruent lexicalization. Nevertheless, Table 13 shows that the dominant pattern of the Mandarin/Southern Min data in Group 5 is insertion rather than congruent lexicalization. This again contradicts the prediction in Table 14.

The predictions of Table 14, at first glance, seem to be confirmed by the results of the other four sets of data, for insertion is the dominant pattern of the Mandarin/Tsou data (i.e., Groups 1 and 2), and congruent lexicalization is the dominant pattern of two sets of Mandarin/Southern Min data (i.e., Groups 3 and 4). However, as mentioned above, the secondary pattern in each group also receives a relatively high score (as shown by the results of Groups 1, 2, and 4). For example, insertion receives a score of 500 in

Group 4, which is very close to the score of 568 received by congruent lexicalization. Since both patterns have very high scores, Muysken's (2000) assumption that a certain extra-linguistic factor favours one of the CS patterns is again problematic.

The last point I want to raise here is that alternation receives relatively low scores throughout all of our data sets if compared with the other two patterns. The results of Deuchar, Muysken and Wang's (forthcoming) Welsh/English<sup>21</sup> data also reflect the same situation. This may be because the Mandarin/Tsou and Mandarin/Southern Min data do not show alternation. However, it is also possible to argue that such a result indicates that the alternation pattern does not exist empirically. Another possibility is that our method of using a clause as the basic unit of analysis is problematic. It has been argued that a clause-based approach assumes the asymmetry between the languages, and therefore the results show a clear preference for insertion and congruent lexicalization.

#### 6.4. Summary

In this chapter, I examined Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach using the Mandarin/Tsou and Mandarin/Southern Min CS data. The results of the tests showed that both models have theoretical problems. The MLF model was applicable to the Mandarin/Tsou data, but not to the Mandarin/Southern Min data. The reason being is that the two major criteria used to identify the ML, namely the morpheme order principle and the system morpheme principle, did not work because of the nature of Chinese languages. Mandarin and Southern Min share most of their syntactic structure and have the same word order

---

<sup>21</sup> In Deuchar, Muysken, and Wang (forthcoming), the scores each pattern received in the English/Welsh data were: Insertion = 710, Alternation = -385, Congruent Lexicalization = 434

(SVO). The type of inflectional morpheme (i.e., the outsider late system morpheme), which is crucial to identifying the ML, is almost absent in these two languages. Hence, neither of the principles was applicable to the Mandarin/Southern Min data sets. To solve the theoretical problem of the MLF model, a revised model was proposed (c.f. section 6.2.3). It was found that the revised MLF model successfully identified the MLs of most of the bilingual clauses in the two sets of Mandarin/Tsou data and the three sets of Mandarin/Southern Min data.

Muysken's (2000) typological approach proposes three different CS patterns, namely insertion, alternation, and congruent lexicalization. His model provides a broader perspective on CS phenomena. It not only looks at intra-clausal CS, but also examines inter-clausal CS and the switches that occur between different turns. One crucial theoretical problem was found when applying his model to my data, namely, there is a lack of a basic unit of analysis. If the unit of analysis is not clear, then it is hard to define what a switch is. If the definition of a switch is unclear, the diagnostic features designed to identify the CS pattern of a given switch cannot be applied. To solve this problem, the key notions from the MLF model (i.e., ML vs. EL opposition) was adopted, and a clause was used as the basic unit of analysis. First, the ML of a given bilingual clause was identified with reference to the four criteria in the revised version of the MLF model. Any elements in the EL were treated as switches. Second, Deuchar's (2004) scoring spreadsheet was used to analyse the two sets of Mandarin/Tsou data and three sets of Mandarin/Southern Min data. This spreadsheet is designed to calculate the overall scores that the three CS patterns receive after applying Muysken's (2000) diagnostic features to an identified switch. I then identified the pattern with the highest score as the dominant pattern of that switch.

The results of my data revealed some other problems with Muysken's (2000) model. First, Muysken's (2000) claim regarding the relationship between CS patterns and linguistic and extra-linguistic factors may be problematic. In none of my five data sets, was an exclusively dominant pattern found. There was always a secondary pattern, which received a score close to the dominant one. In three of my data sets, the differences in scores between the two patterns were not statistically significant. Hence, Muysken's (2000) claim that each CS pattern favours certain linguistic and extra-linguistic factors is questionable. Furthermore, the result of one of the Mandarin/Southern Min (typologically similar and standard/dialect) data sets even showed insertion to be the dominant pattern, rather than congruent lexicalization as predicted by his model. Secondly, alternation received relatively much lower scores across the five sets of data. This may suggest that Muysken's (2000) model (especially the diagnostic features) has a preference for the other two patterns, namely insertion and congruent lexicalization. The same situation holds in the results of Welsh/English CS data (Deuchar, Muysken, & Wang, forthcoming). Hence, the results may also raise the question whether the alternation pattern actually exists.

As mentioned above, Myers-Scotton's (2002a) MLF model and Muysken's (2000) typological approach have theoretical problems. In this chapter, both models were modified and tested with the Mandarin/Tsou and Mandarin/Southern Min data. It was found that the theoretical problem of the original MLF model was solved by the revised version. Muysken's (2000) model, even with modifications, was still problematic. Thus, the revised MLF model will be adopted to analyse the five sets of Mandarin/Tsou and Mandarin/Southern Min data.



## Chapter 7: Re-analysis of the Data Using the Revised MLF Model

### 7.1. Introduction

At the end of the previous chapter, I reported the decision to use the MLF model as the theoretical framework of this study. In section 6.2, it was shown that the two principles of the MLF model (i.e., morpheme order principle and system morpheme principle) could be applied to the Mandarin/Tsou data without any problem. However, there was more difficulty in applying them to the Mandarin/Southern Min data. A revised MLF model using the criteria outlined in Table 7 repeated as Table 15 was found applicable to all Mandarin/Tsou data and most of the Mandarin/Southern Min data. Hence, the revised MLF model shown by Table 15 will be adopted to re-analyse the CS data of these two language pairs.

**Table 15: A Revised Version of the MLF Model**

	<b>Criteria</b>	<b>if not applicable, go to</b>
1	Morpheme Order Principle System Morpheme Principle	(To apply simultaneously) 2
2	Morpheme Counting Principle	
3	Uniform Structure Principle	

In this chapter, the results of applying this revised MLF model to the five sets of Mandarin/Tsou and Mandarin/Southern Min data as well as their possible implications will be discussed. This chapter can be divided into two parts:

(1) **An evaluation of the revised MLF model:** In this part, the revised MLF model will be evaluated by discussing the overall results of its application to the Mandarin/Tsou and Mandarin/Southern Min data. Since the revised MLF model was

not applicable to a small proportion of the Mandarin/Southern Min data, a more detailed discussion of these problematic data will be provided. Then, the results of the analysis of the Mandarin/Tsou data and those of the Mandarin/Southern Min data will be compared.

(2) **Practical implications of the results of the analysis:** For the Mandarin/Tsou data, the results of the analysis will be interpreted from two aspects, namely ML Turnover (Myers-Scotton 1998) and the relationship between EL islands<sup>22</sup> and EL proficiency (Finlayson, Calteaux, & Myers-Scotton 1998). ML turnover refers to the situation in which the original matrix language (i.e., the language which supplies the basic grammatical structure) of a bilingual CP becomes the embedded language and vice versa. The original matrix language is often the endangered language while the original embedded language is the language with higher prestige. Myers-Scotton (1998) argues that if the process of ML turnover is complete, it is very likely that language shift will follow, or in extreme cases language death will occur. With reference to her argument, the results of the analysis of the Mandarin/Tsou data will be examined in order to see if there is any sign of ML turnover. If the answer is yes, I will continue to investigate whether the process of ML turnover is complete or not. A completed ML turnover in the Mandarin/Tsou CS case would mean that the old ML of the majority of the bilingual clauses (presumably Tsou) produced by old Tsou speakers is no longer the ML of those produced by younger Tsou speakers. In other words, the more prestigious language (i.e., Mandarin) would replace Tsou as the new ML. Since Tsou is facing an endangered situation and is suffering from Chinese “invasion” in various levels (e.g., political, socio-economic, educational, and cultural)

---

<sup>22</sup> To recall the discussion in section 3.3.2, “Embedded language islands are full constituents consisting only of Embedded Language morphemes occurring in a bilingual CP that is otherwise framed by the Matrix Language” (Myers-Scotton 2002a, p.139).

(c.f. section 2.5), it is reasonable to hypothesize that an ML turnover process is taking place, or is complete in the Tsou community.

In many minority language communities, it is commonly found that there is a trend of decline in the native languages, either steadily or rapidly, across different generations. For instance, Crystal (2000, p.22) argues that “the knowledge of vocabulary declines, with younger people familiar with only a proportion of the traditional vocabulary known by older people...” according to much research in endangered languages. His argument suggests that younger generations’ proficiency in their native language declines, while their ability to command the more prestigious language in their society increases if compared to older generations. This situation could be a result of education, for the languages with higher prestige are often the only media of teaching in different levels of schools. For example, in South Africa, English is the more prestigious language and is the medium of education in some schools. In their study of Sotho, Zulu, and English CS in Tembisa City of South Africa, Finlayson et al (1998) found that there was a correlation between the informants’ EL proficiency (EL is English) and the frequency of EL islands found in their speech. Although Finlayson et al did not conduct a separate test to measure the informants’ English proficiency, they assumed that the higher education the informants received, the better English proficiency they would have, and vice versa.

The status of Mandarin in Taiwan resembles that of English in South Africa. Both Mandarin and English enjoy higher prestige in politics and in various socio-economic settings, and most importantly, they are the major media of education. According to Finlayson et al. (1998, p. 400), “English is the medium of education from the 4<sup>th</sup> year in Tembisa schools”, while in Taiwan Mandarin is the only medium of education from

the first year of primary school. With reference to Finlayson et al.'s study, a similar analysis will be conducted to investigate whether their findings could be applied to the Mandarin/Tsou data sets or not. Since Mandarin and English have similar status (e.g., both have higher prestige in education), it is hypothesized that younger Tsou people in Group 2 who received higher education in Mandarin would produce more EL islands than would older Tsou people in Group 1 who received lower education.

## 7.2. An Evaluation of the Revised MLF Model

### 7.2.1. Mandarin/Tsou Data

We have two sets of Mandarin/Tsou data which are divided by the subjects' age:

Group 1: Older Tsou People — 55 ~ 70 years old

(They were born between 1935 – 1950)

Group 2: Younger Tsou People — 25~ 40 years old

(They were between 1965 – 1980)

One hundred bilingual Mandarin/Tsou CPs were collected from Group 1. Twenty-one CPs (21%) were found to have Mandarin as their ML, while the other seventy-nine (79%) had Tsou as the ML. In other words, the ML of all the Mandarin/Tsou CS collected from this group can be unambiguously identified. For Group 2, most of the data collected were in Mandarin although all the subjects claimed to be fluent speakers in both Mandarin and Tsou. In the recordings (about six hours), only thirty bilingual CPs were collected. Twenty of them were found to have Mandarin as the ML (67%), and ten to have Tsou as the ML (33%). The details of the data collected from Group 1 and Group 2 are shown in Table 16<sup>23</sup>. As Table 16 shows, the revised MLF

---

<sup>23</sup> The results of Table 16 are the same as those in Table 4.

model was successfully applied to all the Mandarin/Tsou data collected from Group 1 and 2. No problematic data were found.

**Table 16: Results of the Application of the Revised MLF Model to Mandarin/Tsou Data**

	<b>Mandarin= ML</b>	<b>Tsou = ML</b>	<b>Total Number of Bilingual CPs</b>
<b>Group 1</b>	21 (21%)	79 (79%)	100
<b>Group 2</b>	20 (67%)	10 (33%)	30

### 7.2.2. Mandarin/Southern Min Data

The Mandarin/Southern Min data were collected from three sampling groups distinguished by the participants' age and educational level. The details of these three groups are repeated below:

Group 3: Educated below university level/older people — 50 ~ 60 years old

Group 4: University Level/older people — 50 ~ 60 years old

Group 5: University Students — 18 ~ 25 years old

The revised MLF model was applied to one hundred Mandarin/Southern Min bilingual clauses collected from each of the three groups (three hundred in total). The results are shown in Table 8 repeated as Table 17.

**Table 17: The Results of Applying the Revised Version of the MLF Model to Mandarin/Southern Min Data**

	Group 3			Group 4			Group 5		
Matrix Language	Mandarin	Southern Min	Total	Mandarin	Southern Min	Total	Mandarin	Southern Min	Total
Number of Bilingual Clauses	17	79	96	27	69	96	24	74	98
Number of Clauses whose MLs cannot be identified	4		4	4		4	2		2
Total			100			100			100

From Table 17, we can see that the revised MLF model was applicable to more than ninety percent of the total data collected from the three Mandarin/Southern Min groups (Group 3 = 96%; Group 4 = 96%; Group 5 = 98%). The majority of the bilingual clauses across these three groups are found to have Southern Min as the ML (Group 3= 79%; Group 4 = 69%; Group 5 = 74%). This suggests that the sociolinguistic variables (i.e., age and educational level) do not have any influence on the use of ML across the three Mandarin/Southern Min data sets.

*Problematic Data*

As Table 17 shows, some problematic data to which the revised MLF model was not applicable were found. (173a) is one example of the problematic data. Its monolingual Mandarin and Southern Min versions are provided in (173b) and (173c).

(173)a. *yi*                    *nian-ling* **lwe**        **sən**    o? (Mandarin/Southern Min)

according to    age                down    count    part.-ques

‘Do we count the amount of money he can get according to his age?’

(*Italic = Mandarin; bold= Southern Min; underline= language*

unidentifiable)

b. *yi*                    *nian-ling* *qu*                    *suan* *o?* (Mandarin)

according to    age                    go                    count    part.-ques.

‘Do we count the amount of money he can get according to his age?’

c. **ɔ̃ʒiə**                    **ni-gi**                    **lwe**                    **sən**                    **o?** (Southern Min)

according to    age                    down                    count    part.-ques.

‘Do we count the amount of money he can get according to his age?’

To identify the ML of the bilingual clause in (173a), we checked the first two criterion of the revised MLF model shown in Table 15, namely the morpheme order principle (i.e., to check word order) and system morpheme principle (i.e., to check the outsider late system morphemes, such as subject-verb agreement affixes or case markers). If we compare the bilingual clause in (173a) to the monolingual clauses in (173b) and (173c), it is clear that no difference is found in terms of their word order. In addition to that, no late outsider system morpheme is found in (173a). Hence, the first two criterion of the revised MLF model are not applicable to (173a).

Then, we move to the third criterion, namely the morpheme counting principle, which specifies that the language that supplies more morphemes is the ML. However, in (173a) two morphemes (i.e., *according to* ‘*yi*’ and *age* ‘*nian-ling*’) are in Mandarin and two morphemes (i.e., *down* ‘*lwe*’ and *count* ‘*sən*’) are in Southern Min. The language of the underlined sentence final question particle “o” cannot be identified because, as the monolingual examples in (173b) and (173c) shows, it could be either Mandarin or Southern Min. Since Mandarin and Southern Min provide the same amount of morphemes in (173a), the morpheme counting principle is thus not applicable either.

Finally, the last criterion, namely the **uniform structure principle**, was adopted to identify the ML of (173a). This principle argues that in order to keep the uniformity of the grammatical structure of a given bilingual CP, the other two types of system morphemes (i.e., early system morphemes and bridge morphemes) are also likely to come from the ML. To recall the discussion in section 6.2.2.3, the classifiers in Mandarin and Southern Min are identified as early system morphemes, and the possessive marker “de” is identified as a bridge morpheme. In (173a), neither of these two kinds of morphemes is found, and this suggests that the uniform structure principle is not applicable, either. To sum up, the four criteria of the revised MLF model were found not to be applicable to the problematic data. As Table 17 shows, only a very small proportion of the entire three sets of Mandarin/Southern Min CS data were found to be problematic (Group 3= 4%; Group 4 = 4%; Group 5= 2%).

### 7.2.3. Mandarin/Tsou vs. Mandarin/Southern Min data: A Comparison of the Results

Table 17 suggests that in the Mandarin/Tsou data sets, the use of the ML is influenced by the subjects' age. However, in the Mandarin/Southern Min data sets, the influence of extra-linguistic factors (i.e., age and educational level) is not apparent. Both Southern Min and Tsou suffered from oppression as a result of the monolingual Mandarin policy implemented in the 1980s and 1990s. Why does the outcome of their contact with Mandarin then differ? From an extra-linguistic perspective, such an outcome may be because of the size of the population. 73.2 % of the entire population of Taiwan (about twenty-three million) are Southern Min (also known as Southern Min), while the entire population of Tsou is only about six thousand. Hence, the oppression from the strict language policy may cause much greater damages to the Tsou speaking community (e.g., socio-economic disadvantages and the rapid loss of



its speakers), and posed serious danger to its survival. Southern Min however accounts for the great majority of the entire population of Taiwan, so the socio-economic disadvantages were not so apparent. Moreover, the change in the socio-political environment may also influence our results. After the political victory of the DPP party (c.f. section 2.6), the status of Southern Min has been raised while Tsou is still remaining as inferior position.

### 7.3. Practical Implications of the Results of Analysis

No distinction between the use of ML and the different sociolinguistic variables (i.e., age and educational level) was found in the results of the analysis of the three Mandarin/Southern Min data sets. Hence, the focus of this section will be interpreting the results of analysis obtained from the two Mandarin/Tsou data sets.

#### 7.3.1. The Matrix Language (ML) Turnover Hypothesis

##### 7.3.1.1. Main Arguments

Language contact often starts with some degree of lexical borrowing, and then bilingualism or multilingualism follows. In other words, the speakers of the languages in contact become bilingual or multilingual, and are able to produce CS utterances. However, it is very common that one of the languages in contact becomes more dominant because of different political or socio-economic powers behind it. According to Myers-Scotton (2002a, p. 52), “wherever there is bilingualism there is always a power differential between the languages involved- simply because access to sources of power (e.g., high-level jobs, educational facilities, or governmental services) are not equally distributed”. Therefore, as Wurm (1991) argues, speakers of the minority language may abandon (either gradually or rapidly) their native language because of the political and socio-economic benefits associated with the knowledge of

the more prestigious language. Then, stable bilingualism or multilingualism declines, and the more prestigious language may replace the minority language, and become the major medium of communication. This implies language shift. In some extreme cases, the minority language completely loses its speakers and dies. This kind of situation is commonly found in many minority language and endangered language communities.

In the continuum of language contact, CS often comes before language shift or language death, for the speakers need to have sufficient knowledge of two languages in order to switch codes. In some language communities in which the linguistic situation is comparatively stable (i.e., two or more languages have more or less the same status), CS often remains as a common phenomenon, and the survival of these languages is not a problem. However, as mentioned in the previous paragraph, the speakers of an endangered language would give up their native language because of its politically or socio-economically disadvantageous status. In this kind of circumstance, as Crystal (2000) argues, the decline of bilingualism or multilingualism would be very quick. Therefore, CS in an endangered language community would be treated very differently because it is often seen as a sign toward language shift or language death, though this is not always the case. The question is: how can one predict whether a given endangered language will survive or die if CS has become very frequent in that language community?

To answer this question, the Matrix Language (ML) Turnover Hypothesis proposed by Myers-Scotton (1998) needs to be introduced. She argues that “in communities where there is widespread CS or convergence within a CP and where there is a dramatic shake-up in the socio-political balance in favour of the ‘infringing’ language, an ML turnover will result” (p.300). The definition of ML turnover is provided below:

“An ML turnover means that the main language which had structured constituents becomes the minor or Embedded Language (EL); in turn, the language which had been the minor language regarding structure becomes the ML. The result is that grammatical structuring in CPs showing CS is now the task of the new ML.”

(Myers-Scotton 1998, p.299)

This is a description of a situation in which the original matrix language (usually a minority or endangered language), which provides the morphosyntactic structure of a bilingual CP, becomes the embedded language. She goes on to suggest three possible scenarios of ML turnover: First, the original ML (i.e., the minority language) remains as the ML, but some degree of structural borrowing from the more prestigious language is found. She argues that examples of structural borrowing could be an innovation/borrowing of a new linguistic element or an omission of the tense/aspect markers. No example of innovation or borrowing is found in the two Mandarin/Tsou data sets. An invented example would be the borrowing of the classifiers of Mandarin into Tsou, as shown in (174). However, I did find some instances of an omission of the tense/aspect in our corpus, as shown in (175).

(174)a. *na ge xiao-hai*

that class child

‘that child’ (*Mandarin*)

b. *oko eni*

child that

‘that child’ (*Tsou*)

c. oko *ge* eni

child class that

‘that child’ (invented example)

d. mo eobako da Basuya oko *ge* eni.

tense2,Agent hit Obl1 name child class that

‘That child hit Basuya.’ (invented Mandarin/Tsou clause)

As (174a) and (174b) show, the syntactic construction of a classifier phrase in Mandarin is: **demonstrative (numeral) + classifier + noun**, and that of the equivalent expression in Tsou is: **Noun + demonstrative**. Tsou does not have classifiers. However, it is possible for Tsou speakers to add classifiers between the demonstrative and noun, as shown by (174c). Because of long-term contact with Mandarin, the addition of classifiers may become fixed, and the bilingual clause like (174d) may be frequently found in Tsou speakers’ daily conversation.

In addition to the bilingual examples in (174), structural borrowing can also possible be found in a monolingual sentence. Consider the examples in (175) below:

(175)a. bonə do fou ( ?o Basuya)

eat Obl2 meat Nom4 name

‘(Basuya) ate meat.’

(normal print = a monolingual Tsou clause collected from the recordings)

b. *mo* *bonə do fou* ( ?o *Basuya*)

tense2Agent eat Obl2 meat Nom4 name

‘(Basuya) ate meat.’

(bold and italic = a grammatically correct version)

(175a) is the monolingual Tsou clause collected from the recordings and (175b) is its grammatical correct version. It is obligatory to mark tense in Tsou, as shown in (175b). However, it is clear that in (175a) the past tense and agent thematic role marker “mo” is omitted. Since Mandarin is a language which does not mark tense. Hence, it is likely that the omission of the tense marker in (175a) is evidence of structural influence from Mandarin. Myers-Scotton (1998) argues that if instances of structural borrowing, as shown in (174d) or (175a), were found, it could be a signal of an early stage of ML turnover.

The second scenario she suggests is the situation where a composite ML is fossilised as the major medium of communication. According to Myers-Scotton (2002a, p.22), “a composite ML is an abstract frame composed of grammatical projections from more than one variety.” In the case of Mandarin/Tsou CS, for instance, a composite ML, would have Mandarin word order and Tsou late outsider system morphemes (e.g., case markers). This is illustrated by the bilingual Mandarin/Tsou example in (176a). For comparative purposes, the monolingual Mandarin version of this clause is given in (176b). Unfortunately, my translator did not provide the monolingual Tsou version in the transcription. Instead of an actual monolingual clause, we provide a possible syntactic construction in (176c) with reference to the basic VOS word order of Tsou.

- (176)a. *ne jia-yi-da-xue, botgonə ?o ma-jiao-jiao-sho do shi*  
in Jia-yi university many Nom4 plur.-professors all cop.  
*yi-xie zhi-shi-fen-zi.* (Mandarin/Tsou)  
some well-educated people

‘Many professors in Jiy-yi University are well-educated people.’

(normal print = Tsou; *Italic* =Mandarin)

- b. *zai* *jia-yi-da-xue*, *hen-duo* *jiao-sho* *do shi* *yi-xie*  
in Jia-yi university many professor all cop. some  
*zhi-shi-fen-zi*.

well-educated people (*Mandarin*)

‘Many professors in Jiy-yi University are well-educated people.’

- c. *ne* *jia-yi-da-xue*, ..... *botgonə ?o moyoyohəgə tsi* Tsou.  
in Jia-yi university many Nom4 intellectual gen. People  
(a possible syntactic construction of the monolingual Tsou version)

If we compare (176a) to (176b) and (176c), it is clear that (176a) follows Mandarin word order. That is, the subject “many professors” occurs right after the prepositional phrase “in jia-yi university” as shown by (176b), but not at the clause final position as shown by (176c). According to the morpheme order principle, this would point to Mandarin as the ML. However, there is also a presence of the Tsou nominative case marker “*?o*”, locative case marker “*ne*”, and genitive case marker “*tsi*” in (176a). As case markers are treated as late outsider system morphemes, this would point to Tsou as the ML according to the system morpheme principle. Since the basic grammatical structure is supplied by both languages, it is likely that the ML of (176a) is a composite one. However, this is not suggesting that a composite ML is the main medium of communication in the Tsou community. In the total number of one hundred and thirty Mandarin/Tsou bilingual clauses, (176a) is the only one that has a composite ML. Of course, it could be argued that such an example is a speaker’s error because of its very limited occurrence in the Mandarin/Tsou data sets. What this example is likely to represent will be discussed later. The point here is simply to

describe what a composite ML is like.

Note that in the first scenario the minority language remains as the ML in the CS utterances produced by its speakers, though with some structural borrowings from the more prestigious language. In the second scenario, both the original minority language and the more prestigious language build up the basic grammatical structure of the CS utterances. Hence, these two scenarios only show partial progression forwards an ML turnover. The third scenario is the situation where the process of an ML turnover goes to completion. Under this scenario the more prestigious language replaces the minority language (i.e., the old ML) and becomes the new ML of the CS utterances produced by the speakers of the language community in question. Myers-Scotton (1998) argues that the completion of an ML turnover is a sign of language shift and is found in most of the endangered language communities.

The Tsou community has long been a minority group during the Japanese colonisation (1895 – 1945) and the rule of the Chinese Nationalist Party (1945 – 2000). Tsou people lived in geographically remote areas (mainly in high mountains), and were very cooperative toward the rule of Japan. The Japanese government implemented a rather moderate policy toward the Tsou people, and did not interfere too much with the traditional administrative and socio-political system of the Tsou people. The Chinese Nationalist Party (KMT party), at the early stage of its rule, adopted a similar policy until the late 1960s. Hence, Tsou people were able to keep their original culture and life style. However, in 1970s and 1980s, there was a sudden shift in the KMT Party socio-political policy toward all minority ethnic groups in Taiwan because of its

diplomatic failures<sup>24</sup> in international politics (c.f. Chapter 2 for details). A discriminatory socio-political policy (e.g., a strict national language policy) was implemented in order to consolidate the national identity. As a result, the socio-economic status of all minority social groups, including Southern Min, Hakka, and the Austronesian indigenous tribes, severely declined. Such a situation matches the context of the occurrence of an ML turnover in which there is a dramatic shake-up in the socio-political balance in favour of the more prestigious language (Myers-Scotton 1998). Hence, the following discussion will investigate whether the process of an ML turnover is taking place in the Tsou community, and if yes, which scenario it follows.

#### 7.3.1.2. An ML Turnover in Mandarin/Tsou CS

The results in Table 16 indicate a clear change of the ML of most of the bilingual Mandarin/Tsou clauses collected from these two age groups (Group 1 and 2); that is, changing from Tsou to Mandarin. In addition to the quantitative results shown in Table 16, a more qualitative analysis is provided below to illustrate how the change of ML is taking place. Consider the bilingual CP collected from Group 1 (older people) in (177a), and its monolingual Tsou and Mandarin versions in example (177b) and (177c).

- (177)a. mo            *que-shao mo-lian* ?e    iu?faoinana    maitan?e.  
 tense2,Agent lack      training Nom1 young people now  
 ‘Now, young people don’t have enough training.’  
 (normal print = Tsou; Italic =Mandarin)

<sup>24</sup> In 1973, the US and other major countries such as UK and France recognised the communist People’s Republic of China (PRC) as a member state of the United Nation, and terminated their diplomatic relations with the Republic of China (ROC) in Taiwan led by the Nationalist KMT party.



b. mo            loeamamiya    ?e            iu?fafoinana    maitan?e (Tsou)  
 tense2,Agent lack training    Nom1    young people    now  
 ‘Now, young people don’t have enough training.’

c. *xian-zai    nian-qing-ren    que-shao    mo-lian.* (Mandarin)  
 Now            young people    lack            training  
 ‘Now, young people don’t have enough training.’

Tsou is identified as the ML of (177a) by using the two criteria of the original MLF model, namely the morpheme order principle and system morpheme principle. The word order of (177a) is first examined. In (177a), the subject *iu?fafoinana* ‘young people’ occurs after the VP (i.e. *que-shao mo-lian* ‘lack training’), which is the same as the monolingual Tsou example in (177b). If the ML of (177a) were Mandarin, we would expect the subject to occur before this VP as shown by the monolingual Mandarin CP in (177c). Moreover, if we compare the adverbial *maitan?e* ‘now’ in (177a) and (177b), both of them occur at the end of the clause. This again shows Tsou word order. If the ML of (177a) were Mandarin, then this adverbial would occur at the beginning of the clause as shown by the monolingual Mandarin example in (177c). This evidence suggests that (177a) has Tsou word order. Secondly, there is a Tsou nominative case marker *?e* in (177a). According to Myers-Scotton (2002a, p.76), “case affixes” are outsider late system morphemes, which are crucial to identify the ML. Since this kind of morpheme is absent in Mandarin, the presence of the nominative case marker *?e* suggests that the ML of (177a) is Tsou. Both criteria in the MLF model point in the same direction, and Tsou is identified as the ML of (177a). The results in Table 16 show that most of the bilingual Mandarin/Tsou CPs (i.e., 79% as indicated by Table 16) collected from Group 1 (older Tsou people) were like the example in (177a) in which Tsou provides the basic grammatical structure (therefore the ML).

However, if we look at the data collected from Group 2 (younger Tsou people), the situation is entirely different. Consider the bilingual CP in (178a) and its monolingual versions in Tsou and Mandarin:

(178) a. iu?fafoinana *de* *xiang-fa* <sub>subject</sub> *bu* *yi-yiang*.

Young people poss. thought neg. same

‘Young people’s thought was different.’

(normal print = Tsou; Italic =Mandarin)

b. mo a?hto mäh?mæskə ?e totohəgə

tense2,Agent neg. resemble Nom1 thought

da iu?fafoinana <sub>subject</sub>(Tsou)

Obl1 young people

‘Young people’s thought were different.’

c. nian-qing-ren *de* *xiang-fa* <sub>subject</sub> *bu* *yi-yiang*. (Mandarin)

Young people poss. thought neg. same

‘Young people’s thought were different.’

Mandarin is identified as the ML of the bilingual CP in (178a) for two reasons. First, (178a) is in Mandarin word order. The underlined subject iu?fafoinana de xiang-fa ‘young people’s thought’ occurs at the very beginning of the clause which is the same as that of the monolingual Mandarin example in (178c). If Tsou were the ML, we would expect the subject to occur at the clause final position as in (178b).

Second, there is no outsider late system morpheme (i.e. the case marker in Tsou) in (178a), which is a characteristic of Mandarin. If Tsou were the ML, we would expect

to find nominative case marker ‘?e’ in (178a). As shown in Table 16, sixty-seven percent of the bilingual CPs collected from this group are like (178a) and have Mandarin as their ML. If we make a comparison between (177a) and (178a), the occurrence of the ML turnover in the Mandarin/Tsou CS data collected from two generations becomes clearer. That is, the more prestigious language (i.e., Mandarin) has replaced the old ML (i.e., Tsou) as the new ML of most of the CS utterances (67%) produced by young Tsou speakers. Hence, I argue that the process of ML turnover has gone to completion (namely the third scenario proposed by Myers-Scotton (1998)), across two different generations in current Tsou community.

Such a claim is further supported by the fact that most of the utterances produced by Group 2 (younger Tsou speakers) were merely in Mandarin. Only thirty bilingual Mandarin/Tsou CS data were collected in approximately six-hour recordings. Moreover, during the period of my data collection in the Tsou villages, some observations of the language use of even younger generations (i.e., teenagers and children) were conducted. I found that CS utterances were very rare and Mandarin seemed to be used as the only medium of communication by these two generations. This suggests that language shift is taking place in these youngest generations and confirms Myers-Scotton’s argument that a completion of an ML turnover is often followed by language shift. Furthermore, according to Myers-Scotton (1998,p.224), “language shift also involves language death when all speakers of language X abandon that language”. As the politically and socio-economically inferior situation of Tsou is hardly improved, it is likely that language shift will continue, and the survival of Tsou will be problematic.

It could be argued that the only composite CS example shown by (175) is merely a

speaker error. However, it is also possible that the very limited number of composite ML we found in the two Mandarin/Tsou data sets is caused by the gap between the subjects' ages in Group 1 and 2. The age of the subjects in Group 1 ranges from fifty-five to seventy years old, while that of the subjects in Group 2 ranges from twenty-five to forty-five years old. Hence, there is a gap of ten years between the younger in Group 2 and the older in Group 1. It is possible that a higher number of bilingual clauses with a composite ML would have been found if we had collected our data from subjects aged between forty and fifty-five.

### 7.3.2. EL (Embedded Language) Islands and EL Proficiency

Crystal (2000, p.79) argues that in many endangered language communities, “the younger generation becomes increasingly proficient in the new language, identifying more with it, and finding their first language less relevant to their new needs”. It is very likely that such a situation is taking place in the Tsou community. In the previous section, I argued that there was a ML turnover in the Tsou community if comparing the Mandarin/Tsou data collected from Group 1 (older Tsou people) and Group 2 (younger Tsou people). In other words, Mandarin has replaced Tsou as the ML, which provides the grammatical structure of almost all the bilingual Mandarin/Tsou CS data collected from Group 2.

Moreover, the number of the bilingual clauses collected from Group 2 was very limited, and Mandarin was used almost exclusively as the medium of communication in the recordings of this group. It may be hypothesised that the ML turnover found in the two Mandarin/Tsou data sets may be attributed to the decline of young people's proficiency in Tsou, and therefore they have less ability to produce CS utterances with Tsou as the ML. To test such a hypothesis, one needs to evaluate the young Tsou

people's proficiency in Tsou, and see if there is any correlation between their proficiency in Tsou and their use of ML in CS utterances. Unfortunately, no relevant test or examination was conducted to evaluate the subjects' proficiency in Tsou in my fieldwork. However, in their CS study of three languages spoken in South Africa (i.e., Sotho, Zulu, and English), Finlayson, Calteaux, and Myers-Scotton (1998) look at the relationship between language proficiency and EL islands<sup>25</sup> with reference to the major arguments of the MLF model. Their perspective may be useful in terms of understanding the difference between the Tsou proficiency of two different generations.

Finlayson et al. firstly divided their subjects into two different groups based on their educational level. They randomly selected eighty-eight lines of transcribed speech collected from each group, including both monolingual and bilingual sentences. Then, they conducted a quantitative analysis of their data with several goals. One goal that is relevant to this section was to compare the preferred types of CS constituents in the CS utterances produced by these two groups. They found that there was a correlation between the frequency of EL islands and proficiency measured by level of education. They found that Zulu and Sotho were mainly the MLs while English was the EL across all the data produced by these two groups. One of the most striking findings of their study suggested that the more educated subjects produced more EL islands, while the less educated subjects produced more singly-occurring EL forms. It is important to note that they assume, without separate measurement of proficiency, that the higher level of education the subjects received, the higher level of proficiency in English they have, and vice versa. Hence, their finding shows that “embedded

---

<sup>25</sup> To recall the discussion in section 4.4.2.2, “Embedded language islands are full constituents consisting only of Embedded Language morphemes occurring in a bilingual CP that is otherwise framed by the Matrix Language” (Myers-Scotton 2002, p.139).

language proficiency correlates positively with the use of embedded language islands as opposed to singly-occurring lexemes” (p.406).

As a result of their findings, I decided to examine the frequency of EL islands and singly-occurring lexemes in the two sets of Mandarin/Tsou data, and test their prediction that the frequency of EL islands found in CS utterances would follow the differences in Tsou proficiency between our Group 1 (older Tsou people) and Group 2 (younger Tsou people). As mentioned earlier, most of the subjects in Group 1 received some levels of primary school education in Japanese while those in Group 2 received high school education. The details of these two groups are repeated below:

Group 1: Old Tsou people, primary school education level (in Japanese) /did not receive any education

Group 2: Young Tsou people, high school education level (in Mandarin)

Following Finlayson et al.’s (1998) assumption that the higher level of education the subjects received, the higher level of proficiency in English they have, it is assumed that the higher education Tsou people received, the higher level of proficiency in Mandarin they have. The reason is that English is the language with more prestige in South Africa and is the official language as well as the medium of education. The status of Mandarin in Taiwan is similar to that of English in South Africa. Here, I want to investigate whether there is any correlation between the proficiency in Mandarin and the frequency of EL islands (i.e., full constituents in Mandarin) in the bilingual clauses produced by Group 1 and Group 2, which have Mandarin as their ML.

First, the bilingual CPs whose ML is Mandarin were examined and all the elements in

the EL (i.e., Tsou) in these clauses were identified to see if they were singly-occurring lexemes or full constituents. Second, the total number of EL islands and singly-occurring lexemes found for each group was compared. With reference to the finding of Finlayson et al, we would expect the group that produced more EL islands (i.e., full constituents in Tsou) to have better proficiency in Tsou. The procedure followed in my analysis is illustrated by the examples in (179). (179a) is the bilingual Mandarin/Tsou clause, and (179b) and (179c) are its bilingual versions in Tsou and Mandarin.

- (179)a. *zəmoi wen-dao ren de wei-dao hui pao a.*  
 bear smell human poss. smell then run away part.-inform.  
 ‘If bears smell human, they will run away.’
- b. *lazəʔo pəkako homo elə da Tsou ʔe zəmoi.*  
 only run away if smell Obl1 human Nom1 bear  
 ‘If bears smell human, they will run away.’
- c. *xong wen-dao ren de wei-dao hui pao a.*  
 bear smell human poss. smell then run away part.-inform.  
 ‘If bears smell human, they will run away.’  
 (normal print = Tsou; *Italic* =Mandarin)

Mandarin is identified as the ML of (179a), for it has Mandarin SVO word order, but not Tsou VOS word order. The subject *zəmoi* ‘bear’ of (179a) occurs at the clausal initial position, which is the same as that of the monolingual Mandarin clause in (179c). Moreover, no outsider late system morphemes (i.e., the nominative case marker *ʔe*, and the oblique case marker *da* shown by the monolingual Tsou example in (179b)), are found in (179a). Hence, Mandarin is identified as the ML. Then, the

EL element in this bilingual clause was examined to see whether it is a singly-occurring lexeme or a full constituent. It is obvious that the Tsou word *zəmoi* ‘bear’ is a singly-occurring lexeme.

Following the procedure illustrated by the examples in (179), twenty-one bilingual CPs from Group 1 and twenty from Group 2 were analysed. In the twenty-one bilingual CPs collected from Group 1, it was found that twenty-seven Tsou (EL) elements in which seven (26%) were full constituents (and therefore the EL islands) while the other twenty (74%) were singly-occurring lexemes. For Group 2, it was found a total number of twenty-three Tsou (EL) elements in which four (17%) were EL islands while the remaining nineteen (83%) were singly-occurring lexemes. All these results are shown in Table 18.

**Table 18: The Number and Percentage of EL islands vs. Singly-occurring Lexemes in Mandarin/Tsou CPs with Mandarin as ML**

	Total Number of bilingual CPs with Mandarin as ML	Total Number of EL elements in the bilingual CPs with Mandarin as ML	Full Constituents (EL islands) in the bilingual CPs with Mandarin as ML	Singly-occurring Lexemes in the bilingual CPs with Mandarin as ML
<b>Group 1</b>	21	27	7 (26%)	20 (74%)
<b>Group 2</b>	20	23	4 (17%)	19 (83%)

Chi-square was also used to examine whether the difference in the results for Group 1 and Group 2 are statistically significant or not. The results are shown in Table 19.



**Table 19: Nature of EL elements, Group 1 vs. Group 2: EL islands vs. Singly-occurring Lexemes in Bilingual CPs with Mandarin as the ML**

---

	Full Constituents (EL islands)	Singly-occurring Lexemes in the	<b>Total</b>
	in the bilingual CPs with	bilingual CPs with Mandarin as	
	Mandarin as ML	ML	
<b>Group 1 (Old)</b>	7	20	27
<b>Group 2 (Young)</b>	4	19	23
<b>Total</b>	11	39	50

---

$X^2 = 0.169$ ,  $df=1$  (not significant)

Table 19 suggests that the obtained value of  $X^2$  (0.169) does not reach the level of significance (critical value= 3.841 which is significant at the 0.05 level). Hence, it may be concluded that there is no clear relation between the proficiency in Tsou and the frequency of EL islands (i.e., full constituents in Tsou) in the bilingual clauses produced by Group 1 and Group 2.

A similar analysis was also conducted to examine the bilingual clauses which have Tsou as their ML. This time, it was aimed to investigate whether there is any correlation between the proficiency in Mandarin and the frequency of EL islands (i.e., full constituents in Mandarin in the bilingual clauses produced by Group 1 and Group 2. It was assumed that the more EL (Mandarin) islands to be found in my data, the higher level of proficiency the subjects would have in Mandarin. The examples in (177), repeated as (180), will be used to illustrate the detailed procedures of the analysis.

(180)a. mo            *que-shao*   *mo-lian*   ?e     *iu?fafoinana*   *maitan?e*.  
 tense2,Agent   lack            training   Nom1   young people   now  
 ‘Now, young people don’t have enough training.’  
 (normal print = Tsou; Italic =Mandarin)

b. mo            *loeamamiya*   ?e            *iu?fafoinana*   *maitan?e* (Tsou)  
 tense2,Agent   lack training   Nom1   young people   now  
 ‘Now, young people don’t have enough training.’

c. *xian-zai*     *nian-qing-ren*   *que-shao*     *mo-lian*. (Mandarin)  
 Now            young people   lack            training  
 ‘Now, young people don’t have enough training.’

(180a) is an example of Mandarin/Tsou CS. Tsou is identified as the ML of (180a), for it has Tsou VOs word order but not Mandarin SVO word order. The subject *iu?fafoinana zəmoi* ‘young people’ of (180a) as well as the adverbial *maitan?e* ‘now’ occur at the clausal final position, which are the same as that of the monolingual Tsou clause in (180b). Moreover, an outsider late system morpheme, namely the Tsou nominative case marker “?e” can be found in (180a). Hence, Tsou is identified as the ML. If Mandarin were the ML, we would expect the subject *iu?fafoinana zəmoi* ‘young people’ to occur at the clause initial position, and no case markers would occur.

The EL element, namely the Mandarin elements *que-shao mo-lian* ‘lack training’ in this bilingual clause, is then examined in order to see if it is a singly-occurring lexeme or an EL island. It is obvious that these two Mandarin elements constitute a verb phrase, and therefore is an EL island. Following the same procedures, eighty-nine bilingual CPs (including seventy-nine from Group 1 and ten from Group 2) whose ML is Tsou were examined. In the seventy-nine bilingual CPs collected from Group 1, It was found that one hundred and twenty-nine Mandarin (EL) elements in which

thirty-two (24.8%) were full constituents (and therefore the EL islands) while the other ninety-seven (75.2%) were singly-occurring lexemes. In the ten bilingual CPs collected from Group 2, it was found that a total of twenty-one Mandarin (EL) elements in which ten (47.6%) were full constituents and the remaining eleven (52.4%) were singly-occurring lexemes. All these figures are shown by Table 20.

**Table 20: The Number and Percentage of EL Islands vs. Singly-occurring Lexemes in Mandarin/Tsou CPs with Tsou as ML**

	<b>Total Number of bilingual CPs with Tsou as ML</b>	<b>Total Number of EL elements in the bilingual CPs with Tsou as ML</b>	<b>Full Constituents (EL islands) in the bilingual CPs with Tsou as ML</b>	<b>Singly-occurring Lexemes in the bilingual CPs with Tsou as ML</b>
<b>Group 1</b>	79	129	32 (24.8%)	97 (75.2%)
<b>Group 2</b>	10	21	10 (47.6%)	11 (52.4%)

Again, a chi-square test was conducted to examine whether the difference in the results for Group 1 and Group 2 are statistically significant or not. The results are shown in Table 21.

**Table 21: Nature of EL Elements, Group 1 vs. Group 2: EL Islands vs. Singly-occurring Lexemes in Bilingual CPs with Tsou as the ML**

	Full Constituents (EL islands) in the bilingual CPs with Tsou as ML	Singly-occurring Lexemes in the bilingual CPs with Tsou as ML	Total
<b>Group 1</b>	32	97	129
<b>Group 2</b>	10	11	21
<b>Total</b>	42	108	150

$X^2 = 3.599, df = 1$

Although Table 21 suggests that the obtained value of  $X^2$  (3.5999) is not statistically significant, but it is very close to reach the level of significance (the critical value of 3.841 which is significant at the 0.05 level). Such a result may be caused by the very limited number of bilingual CPs collected from Group 2 (younger Tsou people). The result might be significant if more data could have been obtained. Hence, the results in Table 20 and 21 could be interpreted more positively. In other words, it is likely that there is a correlation between the proficiency in Mandarin and the frequency of EL islands (i.e., full constituents in Mandarin) in the bilingual clauses produced by Group 1 and Group 2.

To sum up, in the analysis of the bilingual clauses with Mandarin as the ML, the relation between the frequency of EL (Tsou) islands and EL proficiency was not apparent. Conversely, the frequency of EL (Mandarin) islands and EL proficiency was found nearly significant in those bilingual clauses with Tsou as the ML. A possible

explanation for such a contrast is that Mandarin (similar to English in South Africa) is used as the medium of education in Taiwan. The subjects in Group 2 (younger Tsou people) received higher education than those in Group 1 (older Tsou people). With reference to Finlayson et al's argument, they have higher proficiency in Mandarin, and are able to produce more EL islands in their CS utterances. Tsou, however, does not share the same status as Mandarin. Hence, the relationship between Tsou proficiency and the frequency of Tsou EL islands is not obvious.

## Chapter 8: Summary and Conclusion

### 8.1. Summary

This thesis had two major aims. The first was the attempt to answer the question: is there a universally applicable CS model? The second aim of this thesis was a more practical one, namely to predict the future of Tsou, an endangered language.

Myers-Scotton's (2002a) MLF model, a prominent model, is claimed to be universally applicable and was considered in attempting to answer the first question. It has been successfully applied to different language pairs which are typologically very different in terms of their word order. In this thesis, the MLF model was tested with two previously untested language pairs, namely Mandarin/Tsou and Mandarin/Southern Min. The results of the test showed that the MLF model could unambiguously identify the MLs of all the Mandarin/Tsou bilingual clauses. However, it was not applicable to the Mandarin/Southern Min data, for Mandarin and Southern Min have exactly the same word order, and the type of morpheme (i.e., the outsider late system morpheme), which is crucial to identify the ML, is absent in these two languages. This suggests that the original MLF model could work well under two conditions. First, the participating languages in a given CS utterance should be typologically very different in terms of their word order. Second, at least one participating language should have a comparatively greater amount of inflectional morphology (implying that it is either an inflectional or an agglutinating language). Both Mandarin and Southern Min are isolating languages, hence, they have only a very limited amount of inflectional morphology, and they share most of their syntactic structure. Therefore, the two important principles of the MLF model (i.e., the morpheme order principle and the system morpheme principle) do not apply.

For several reasons the other morphosyntactic CS model, namely Muysken's (2000) typological approach, was also tested in this thesis. First, because this approach is also a prominent one, but it contrasts with the MLF model in terms of adopting a more comprehensive perspective on CS phenomena. He argues that the MLF model only looks at one possible pattern of CS, namely insertion. He suggests that other CS patterns, namely alternation and congruent lexicalization that should also be taken into account in order to give a fuller picture. The second reason for testing Muysken's model was that the MLF model was only applicable to the Mandarin/Tsou data, but did not apply to the Mandarin/Southern Min data. Therefore, it was necessary to find another model which accounted for both language pairs. Third, the basic unit of analysis of the MLF model is a bilingual CP, and therefore it only allows us to examine intra-clausal switches. The typological approach, however, not only looks at intra-clausal switches, but also inter-clausal switches and the switches which occur between different turns.

However, one fundamental problem was found when testing the typological approach with my two language pairs, namely, the basic unit of analysis is not clear. Hence, it was difficult to decide what counts as a switch. For this reason, it was then impossible to apply those diagnostic features that Muysken proposed to identify the CS pattern of a given switch.

The results of my analysis showed that both the MLF model and the typological approach are problematic. However, the basic unit of analysis of the MLF model is very clear, whereas that of the typological approach is not. I decided to modify the MLF model referring to an earlier version of the model which had included the

morpheme counting principle and an associated principle, the uniform structure principle when processing the Mandarin/Southern Min data. The revised model is shown in Table 7 repeated as Table 22 below.

**Table 22: A Revised Version of the MLF Model**

	<b>Criteria</b>		<b>if not applicable, go to</b>
1	Morpheme Order Principle	(To apply simultaneously)	2
	System Morpheme Principle		
2	Morpheme Counting Principle		3
3	Uniform Structure Principle		

The four criteria being suggested in the revised MLF model should be applied in a specific order. The two criteria at stage 1 should be applied first to identify the ML of a given bilingual CP. If the ML is not identified, then the second will be used. If the second fails, then we move to the third. This revised MLF model was used to re-analyse the Mandarin/Tsou and Mandarin/Southern Min data. The results of the re-analysis showed that this revised model successfully applied to all the Mandarin/Tsou data and to the great majority of the Mandarin/Southern Min data.

For the two Mandarin/Tsou data sets, most of the bilingual clauses collected from Group 1 (older Tsou people) were found mainly to have Tsou as the ML, while those collected from Group 2 (younger Tsou people) were found predominantly to have Mandarin as the ML. For the Mandarin/Southern Min data sets, Southern Min was found to be the ML of most of the bilingual clauses collected from Groups 3, 4 and 5. In other words, the MLs of the bilingual clauses in the two Mandarin/Tsou data sets were different, but that was not so for the bilingual clauses in the three Mandarin/Southern Min data sets, which did not differ with respect to the ML. A



sociolinguistic interpretation was provided to explain why the results of the analysis of these two language pairs differed in that way. Both the Tsou and the Southern Min people had suffered oppression from the monolingual Mandarin policy implemented in the 1980s and 1990s. However, because of the socio-political success of the Southern Min speakers and their sheer weight of numbers, the predominant use of Southern Min as an ML shows its relatively healthy status. In contrast, the Tsou community is seriously disadvantaged in terms of its population size, political and socio-economic status. This is reflected in the change of ML from the older to the younger group, which, I shall argue below, indicates its status as an endangered language.

## 8.2. Concluding Remarks

### 8.2.1. The Universal Applicability of the Revised MLF Model

Many CS models attempt to relate CS phenomena to a more general linguistic perspective, namely the distinction between a speaker's competence and performance. Competence refers to the speaker's knowledge of a language while performance refers to how a language is actually used in a given situation (Chomsky 1965). Myer-Scotton's (2002a) MLF model also attempts to address the issues of competence and performance.

The premise of the MLF model is that there is always a distinction between the ML and EL in a given bilingual clause in which the two participating languages do not contribute equally. The basic morphosyntactic structure of a bilingual clause is always supplied by the ML. The 4-M model, an augmented model of the MLF model, clearly distinguishes four different types of morphemes with reference to their accessibility in the language production process. She argues that "the accessibility is part of

competence” while how different types of morphemes are accessed differently in the process of language production is a matter of speaker’s performance (p.27). In other words, by examining the bilingual speech, the MLF model, as Myers-Scotton argues, aims to provide a fuller account of the interaction of the speaker’s competence and performance, and therefore to achieve universal applicability.

The MLF model has unambiguously identified the MLs of the CS data of many different language pairs reported in the current literature. The two grammatical constraints of the MLF model, namely the morpheme order principle and the system morpheme principle have achieved considerable empirical success. It is nevertheless true that the notion of the ML construct as proposed by Myers-Scotton (2002a) has been challenged by some scholars, such as, Boussofara-Omar (2003) and Macswan (2005), and the existence of a ML has been questioned. Some counter-examples have been proposed to argue against the universal applicability of this model. However, as discussed in Chapter 6, those so-called counter-examples were results of misinterpretation of the MLF model.

One language pair, namely the Mandarin/Southern Min CS data in the present study appeared to pose an empirical problem for the MLF model. That is, the morpheme order principle and the system morpheme principle were found not to be applicable to the Mandarin/Southern Min CS data. However, this does not mean the MLF model has to be rejected. In fact, no example which violated the MLF model was found in my Mandarin/Southern Min corpus, which includes three hundred bilingual Mandarin/Southern Min clauses. I therefore argue that the production of the Mandarin/Southern Min CS data still follows the same language production process specified by the MLF model, namely the involvement of the ML vs. EL opposition

and the content vs. system morpheme opposition. The inapplicability of the MLF model was caused by the fact that Mandarin and Southern Min share most of their syntactic structure and neither language has the outsider late system morphemes. Such linguistic features make the role of the ML more implicit, and therefore additional criteria to identify the ML are required.

The revised MLF model this thesis has proposed, as shown in Table 22 above, was able to unambiguously identify the MLs of the great majority of the bilingual Mandarin/Southern Min clauses, and therefore solves the problem of inapplicability that the original MLF model has. Furthermore, this revised model not only preserves the arguments offered by the original MLF model that the production of CS utterances is a result of interaction of a speaker's competence and performance, but also provides a fuller account of CS phenomena in an empirical base. So, what, then, is the answer to the question of whether there is a universally applicable CS model? My answer is "yes". Associated with the empirical success of the original MLF model and with some modifications of its original criteria with reference to empirical problems encountered, I argue that the revised MLF model is a model of such a kind.

#### 8.2.2. The Future of the Endangered Language Tsou

As indicated earlier in this chapter, the second aim of this thesis was a more practical one, namely to predict the future of Tsou, an endangered language. The ML Turnover Hypothesis (Myers-Scotton 1998) was adopted to help predict the future of Tsou. I have shown that most of the bilingual Mandarin/Tsou clauses collected from Group 1 (older Tsou people) had Tsou as the ML, while the majority of those collected from Group 2 (younger Tsou people) were found to have Mandarin as the ML. This meant that a ML turnover process had happened. That argument was further supported by

other evidence. First, most of the utterances produced by Group 2 were in Mandarin. In approximately six hours of recordings of younger Tsou people's conversations, only thirty bilingual Mandarin/Tsou clauses were found; all occurred in one recording of approximately forty minutes. The great majority<sup>26</sup> of the clauses collected from this particular recording were also in Mandarin. Second, it was observed that the even younger generations (including teenagers and children) seemed to use only Mandarin in their conversations. These two factors suggest that language shift is taking place. As Myers-Scotton (1998) argues, the completion of an ML turnover is often followed by language shift. Hence, the two factors mentioned above indirectly support my argument that an ML turnover process has happened. Since an ML turnover has already occurred, and language shift is in progress, it is predicted that the survival of Tsou is in doubt.

### 8.3. Limitations and Suggestions for Future Studies

In Chapter 2, it was mentioned that the citizens of the cities in southern Taiwan (e.g., Tainan city where this study was conducted) are the major forces which support the DPP party. Their main political goal is to react against the regime established by Mandarin Chinese speakers (i.e., the KMT party) and to claim the political independence of Taiwan. Their political position is influenced by various historical factors, such as the Japan's colonization of Taiwan and the KMT government's strict rule. The political position and Southern Min group identity of the subjects may play a role in influencing their language use. Therefore, the great majority of the bilingual CPs collected from Groups 3 to 5 have Southern Min as their ML. It might not be the case if the data had been collected in a northern city like Taipei, the capital of Taiwan,

---

<sup>26</sup> 80% of the clauses (260 out of 325) collected from this particular recording were in Mandarin; 11% were in Tsou (35 out of 325); 9% were bilingual Mandarin/Tsou CPs (30 out of 325).

where, the sampling groups would have a different political orientation. Hence, the results of the analysis of Mandarin/Southern Min CS data may be limited to the southern part of Taiwan. Future research could collect further data from the northern part of Taiwan and compare the results.

For the Mandarin/Tsou data, the amount of CS data collected from younger Tsou people was rather limited because of the limit of time and the influence of natural forces. It is suggested that future research could include more subjects and involve longer recordings to test some of the findings of this thesis. For instance, the sociolinguistic variable of age could be examined in more detail. Data from only two age groups, namely Group 1 (55 – 70 years old) and Group 2 (25 – 40 years old) were collected. Future studies could collect CS data from the missing age range (i.e. 40 – 55 years old) and examine the morphosyntactic structure of those data with reference to the revised MLF model and the ML turnover hypothesis. Furthermore, the quantity of the CS data collected from younger generations (i.e. Group 2) was rather limited. Most of those CS data were found to have Mandarin as the ML. Future researchers could either record the data from the same age group in other Tsou villages or collect CS data from even younger generations, such as teenagers, or children in order to provide data to investigate the endangered situation of the Tsou language.

Regarding the theoretical limitation of the present day, the revised MLF model could unambiguously identify the MLs of the great majority of the Mandarin/Southern Min bilingual clauses (c.f. section 7.2.2), but not all. The MLs of those that I referred to as problematic data were unable to be identified because the two participating languages involving in CS supplied equal numbers of morphemes. I am aware that the application of the morpheme counting principle has a certain degree of arbitrariness,

which may lead to some more general theoretical problems. For instance, one may ask the questions: is the process of calculating the morphemes of the participating languages really involved in the production of CS utterances? If yes, is it related to the speaker's competence in a given language or is it simply a matter of performance? I concede that I have not found answers to these questions.

The morpheme counting principle and the uniform structure principle are included in the revised MLF model mainly for empirical reasons. In Chapter 3, various predominant CS models were evaluated. While some sociolinguistic variables (e.g., age) were adopted to interpret some of the results of the analysis, the CA approach was abandoned because its focus is rather different from the research aims of the present study. Other competing structural models, which do not invoke an ML construct, were rejected because of theoretical or empirical problems.

Poplack's (1980) model was rejected by the reported counter-examples in previous studies and by the Mandarin/Tsou CS data. Macswan's (2005) model was proposed with reference to Chomsky's (1995) latest syntactic theory, namely the minimalist programme, which has been widely applied to explain the general process of language production in human brain. The PF Disjunction Theorem of Macswan's (2005) model predicts that switching of phonological systems between a stem and an affix is prohibited. Since Mandarin and Southern Min are both isolating languages in which the number of free morphemes is very limited, the prediction of the PF Disjunction Theorem then does not apply to the Mandarin/Southern Min CS data. Although the MLF model has the same problem of inapplicability to the Mandarin/Southern Min CS data, that problem was solved by the adoption of additional criteria to identify the ML of the Mandarin/Southern Min CS data. The revised MLF model therefore has

more advantages than other competing models in terms of its empirical success, and this was the major reason for adopting it to analyse the CS data in the two language pairs in this thesis.

## References

- Administrative Office of Taiwan Provincial Government R.O.C. *Taiwan Provincial Government Release 1945 - 1996*. Nan-to: Administrative Office of Taiwan Provincial Government R.O.C
- Auer, Peter (1984). *Bilingual Conversation*. Amsterdam: Benjamins
- Auer, Peter (1998). Introduction: Bilingual Conversation Revisited. In Auer, Peter (ed.), *Codeswitching in Conversation*. London & New York: Routledge
- Auer, Peter (2000). A Conversation Analytic Approach to Codeswitching and Transfer. In Li, Wei (ed.). *The Bilingualism Reader*. London and New York: Routledge, pp. 166 -187
- Auer, Peter (2005). A Postscript: Code-switching and Social Identity. *Journal of Pragmatics* 37: pp. 403-410
- BBC News: [http://news.bbc.co.uk/1/hi/world/asia-pacific/country\\_profiles/1285915.stm](http://news.bbc.co.uk/1/hi/world/asia-pacific/country_profiles/1285915.stm), access on February 22, 2006
- Bellwood, Peter (1991). The Austronesian Dispersal and The Origin of Languages. *Scientific American* 265:1, pp. 88-93
- Blom, Jan-Petter & Gumperz, J. John (2000). Social Meaning in Linguistic Structure: Codeswitching in Norway. In Li Wei (ed). *The Bilingualism Reader*, pp. 111-136. London & New Yor: Routledge
- Blust, Robert (1985). The Austronesian Homeland: A Linguistic Perspective. *Asian Perspective* 26:1, pp. 45-67
- Boussofara-omar, Naima (2003). Revisiting Arabic Diglossic Switching in Light of the MLF Model and Its Sub-models: the 4-M Model and the Abstract Level Model. *Bilingualism: Language and Cognition* 6:1, pp. 33-46.
- Chang, Melody Ya-Yin (1998). *Wh-Constructions and The Problem of Wh-Movement in Tsou*. MA Dissertation. Hsinchu: National Tsing Hua University
- Chang, Melody Ya-yin (2004). *Subjecthood in Tsou Grammar*. PhD Thesis. Hsinchu: National Tsing Hua University
- Chen, Chiou-mei (1989). *A Study on Taiwanese Sentence-Final Particles*. MA Dissertation. Taipei: National Taiwan Normal University
- Chen, Mei-ru (1998). *A Review and The Future of the Language Education Policy in Taiwan* <台灣語言教育政策之回顧與展望>. Kaohsiung: Fu Wen Publishing Company
- Chen, Pei-feng (2006). *The Different Intentions behind of the Semblance of "Douka": The Language Policy, Modernisation and Identity in Taiwan during the Japan-Ruling Period*. Taipei: Mai-tian Publisher Ltd.
- Cho, Taehong (1997). Variations and Universals in VOT: Evidence from 17 Endangered Languages. *University of California Working Papers in Phonetics* 95,



DC. pp. 18-40

- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA MIT Press.
- Chun, Cheng-Cheng (2005). The Role of Language in the Taiwaners' Identity. Papers presented in *The 5<sup>th</sup> International Symposium on Bilingualism* at Universitat Politècnica de Catalunya, Barcelona, Spain.
- Clyne, Michael (2003). *Dynamic of Language Contact*. Cambridge: Cambridge University Press
- Council of The Indigenous People, Taiwan: [www.tacp.gov.tw/INTRO/FMINTRO](http://www.tacp.gov.tw/INTRO/FMINTRO); access on January 23, 2006
- Crystal, David (1997). *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishers.
- Crystal, David (2000). *Language Death*. Cambridge: Cambridge University Press.
- DiSciullo, Anne-Marie, & Muysken, Pieter, & Singh, Rajendra (1986). Code-mixing and government. *Journal of Linguistics* 22, pp. 1-24
- Deuchar, M. (2004). *Minority Language Survival: Code-mixing in Welsh*. Proceedings of the Fourth International Symposium on Bilingualism. Tempe: Arizona State University.
- Deuchar, Margaret (2006). Welsh-English Code-switching and The Matrix Language Frame Model. *Lingua* 116:11, pp.1745-2022
- Deuchar, Margaret, & Pieter, Muysken, & Sung-Lan, Wang (2007). Structured Variation in Code-switching Towards An Empirically-based Typology of Bilingual Speech Patterns. *The International Journal of Bilingualism and Bilingual Education* 10:3, pp. 298-340
- Dyen, Isidore (1956). Language Distribution and Migration Theory. *Language* 32:4, pp. 611-626
- Finlayson, Rosalie, & Calteaux, Karea, & Myers-Scotton, Carol (1998). Orderly Mixing and Accommodation in South African Codeswitching. *Journal of Sociolinguistics* 2:3, pp. 395-420
- Fishman, A. Joshua (1965). Who Speaks What Language to Whom and When? In Li, Wei (ed). *The Bilingualism Reader*. London and New York: Routledge
- Fuller, M. Janet, & Lehnert, Heike (2000). Noun Phrase Structure in German-English Code-switching: Variation in Gender Assignment and Article Use. *The International Journal of Bilingualism* 4:3, pp.399-420
- Gafaranga, Joseph (2005). Demythologising Language Alternation Studies: Conversational Structure vs. Social Structure in Bilingual Interaction. *Journal of Pragmatics* 37: pp. 281-300
- Gumperz, J. John (1982). *Discourse Strategies*. Cambridge: Cambridge University

- Press.
- Heller, Monica (1992). The Politics of Code-switching & Language Choice. In Eastman, M. Carol (ed.), *Codeswitching*. Clevedon, Philadelphia & Adelaide: Multilingual Matters Ltd.
- Hsu, Chin-fa (ed.) (2004). *Selected Documents of the History of Taiwan*. Taipei: Academic Sincia.
- Hu, Jianhua, Pan, Haihua, & Xu, Liejiong (2001). Is There A Finite vs. Nonfinite Distinction in Chinese? *Linguistics* 39:6, pp. 1117-1148.
- Huang, Huie-Ju (2005). Repair in Verb-Initial Languages. *Language and Linguistics* 6:4, Oct. pp.575-597
- Huang, Shuanfan (2000). Language Identity and Conflict: A Taiwanese Study. *International Journal of The Sociology of Language* 143, pp.139-149
- Jake, L. Janice, & Myers-Scotton, Carol, & Gross, Steven (2002). Making A Minimalist Approach to CS work: Adding the Matrix Language. *Bilingualism: Language and Cognition* 5:1, pp.69-71.
- Kamwangamalu, M. Nkonko, & Lee, Cher-Leng (1991). Chinese-English Code-mixing: A Case of Matrix Language Assignment. *World Englishes* 10:3, pp.247-261.
- Kho, Kek-tun (2000). *An Introduction to Taiwanese Language <台灣語概論>*. Taipei: Avanguard Publishing Company
- Ko, Fa-chuan (2001). *Negation in Tsou*. MA Disseration. Hsinchu: National Tsing Hua University.
- Kuo, Yun-Hsuan (2005). *New Dialect Formation: The Case of Taiwanese Mandarin*. PhD thesis. Essex: University of Essex
- Labov, Williams (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Li, Ing. Cherry (1999). *Utterance-Final Particles in Taiwanese: A Discourse-pragmatic Analysis*. Taipei: The Crane Publishing Co. Ltd
- Li, N. Charles, & Thompson, A. Sandra (1981). *Mandarin Chinese*. Berkely, Los Angeles & London: University of California Press Ltd.
- Li, Paul Jen-kuei (1999). *History of Formosan Natives: Linguistic Perspective*. Nanto City: The Reference Committee of Taiwan Province
- Li, Wei (1994). *Three Generations Two Languages One Family*. Clevedon & Philadelphia & Adelaide: Multilingual Matters LTD.
- Li, Wei (2002). "What do you want me to say?" On the Conversation Analysis Approach to Bilingual Interaction. *Language in Society* 31, pp. 159–180
- Li, Wei (2005). How can you tell? Towards a common sense explanation of conversational code-switching. *Journal of Pragmatics* 37, pp. 375-389

- Li, Yafei (1990). On V-V Compounds in Chinese. *Natural Language and Linguistic Theory*: 8, pp. 177-207.
- Liao, Chao-Chih (2000). Changing Dominant Language Use and Ethnic Equality in Taiwan Since 1987. *International Journal of The Sociology of Language* 143, pp. 165-182
- Local Government of The Alishan Region: <http://www.alishan.gov.tw/> (Accessed on: April 15, 2005)
- Lonely Planet: [http://www.lonelyplanet.com/mapshells/north\\_east\\_asia/taiwan/taiwan.htm](http://www.lonelyplanet.com/mapshells/north_east_asia/taiwan/taiwan.htm) (Accessed on April 20, 2005)
- Macswan, Jeff (2000). The Architecture of the Bilingual Language Faculty: Evidence from Intrasentential Code Switching. *Bilingualism: Language and Cognition* 3:1, pp. 37-54
- Macswan, Jeff (2005). Codeswitching and Generative Grammar: A Critique of The MLF Model and Some Remarks on “Modified Minimalism”. *Bilingualism: Language and Cognition* 8:1, pp. 1-22.
- Melchert, H. Craig (1980). Some Aspects of ‘Aspect’ in Mandarin Chinese. *Linguistics* 18, pp. 635-654.
- Muysken, Pieter (2000). *Bilingual Speech*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol (1991). Intersections between Social Motivations and Structural Processing in Codeswitching. *Papers for the Workshop on Constraints, Conditions and Models*, pp. 57-82. Strasbourg: European Science Foundation
- Myers-Scotton, Carol (1993/1997). *Duelling Languages: Grammatical Structure in Code-switching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol & Jake, Janice L. (1995). Matching Lemmas in Bilingual Language Competence and Production Model: Evidence from Intrasentential Code-Switching. *Linguistics* 33, p.991-1024
- Myers-Scotton, Carol (1998). A Way to Dusty Death: The Matrix Language Turnover Hypothesis. In Grenoble, Lenore A, & Lindsay, J. Whaley (eds.). *Endangered Languages*, pp.289 –316. Cambridge: Cambridge University Press
- Myers-Scotton, Carol (2000). Four Types of Morpheme: Evidence from Aphasia, Code-switching, and Second-Language Acquisition. *Linguistics* 38:6, pp.1053-1100
- Myers-Scotton, Carol & Jake, Janice L. (2000). Testing the 4-M Model: An Introduction. *The International Journal of Bilingualism*, 4:1, pp.1-8
- Myers-Scotton, Carol. & Bolonyai, Agnes. (2001). Calculating Speakers: Codeswitching in A Rational Choice Model. *Language in Society* 30, pp.1-28
- Myers-Scotton, Carol (2002a). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford & New York: Oxford University Press
- Myers-Scotton, Carol (2002b). Review of Bilingual Speech: A Typology of

- Code-Mixing. *Language* 78:2, pp. 330-333
- Myers-Scotton, Carol (2004). Research Note and Erratum. *Bilingualism: Language and Cognition* 7:1, pp. 89 -90
- Myers-Scotton, Carol (2006). *Multiple Voices: An Introduction to Bilingualism*. Malden & Oxford: Blackwell Publishing Ltd.
- Nartey, Jonas (1982). Code-switching, Interference or Faddism? Language use among Educated Ghanians. *Anthropological Linguistics* 24:2, pp.183-192.
- Norman, Jerry (1988). *Chinese*. New York: Cambridge University Press.
- Paradis, Johanne, & Nicoladis, Elena, & Genesee, Fred (2000). Early Emergence of Structural Constraints on Code-mixing: Evidence from France-English Bilingual Children. *Bilingualism: Language and Cognition* 3:3, pp. 245 - 261
- Poplack, Shana (1980). Sometimes I'll Start A Sentence in Spanish Y TERMINO EN ESPAÑOL: Toward A Typology of Code-Switching. *Linguistics* 18: pp. 581-618
- Poplack, Shana. & Walker, James A. (2003). Review of Bilingual Speech: A Typology of Code-mixing. *Journal of Linguistics* 39, pp. 678 -682
- Ruhlen, Merritt (1987). *A Guide to the World's Languages*. London: Edward Arnold Publisher Ltd.
- Suemitsu, Kinya (2004). *The History of Taiwan: The Japanese Colonial Period*. Taipei: Zhi-liang Publisher Ltd.
- Sun, Chao-Fen, & Givon, Talmy (1985). On The So-called SOV Word Order in Mandarin Chinese: A Qualified Text Study and Its Implication. *Language* 61:2, pp.329-351
- Tainan City Council: <http://www.tncg.gov.tw/> (Access on April 10, 2005)
- The University of Texas, Austin: [http://www.lib.utexas.edu/maps/cia04/taiwan\\_sm04.gif](http://www.lib.utexas.edu/maps/cia04/taiwan_sm04.gif); Access Date: April 10, 2005)
- Treffers-Daller, Jeanine (1994). *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin & New York: Mouton de Gruyter
- Tse, John Kwock-Ping (2000). Language and A Rising New Identity in Taiwan. *International Journal of The Sociology of Language* 143, pp. 151 -164
- Tsuchida, S. (1976). *Reconstruction of Proto-Tsouic Phonology*. Tokyo: Study of Languages and Cultures of Asia and Africa, Monograph series No.5.
- Tung, Tung-ho (1964). *A Descriptive Study of the Tsou Language, Formosa*. Taipei: Academia Sinicia.
- Wei, Longxing (2001). Lemma Congruence Checking between Languages as An Organizing Principle in Intrasentential Code-switching. *The International Journal of Bilingualism* 5:2. pp.152-173
- Wu, Xing-rong (1938). *The Dairy of Wu Xing-rong*. Taipei: Yuan-jin Publisher Ltd.
- Wurm, Stephen A. (1991) Language Death and Disappearance: Causes and

- Circumstances. In Robins, Robert H. & Uhlenbeck, Eugenius M (eds.). *Endangered Languages*. Oxford & New York: Berg Publishers Limited
- Yang, Xiofang (2000). *The Syntax of Southern Min Dialect in Taiwan* <台灣閩南語語法稿>. 4<sup>th</sup> ed. Taipei: Da An Publisher.
- Yide, Kiwata (1937). *A Review of Japanese Colonial Government Policies in Taiwan* <台灣治績志>. Taipei: Ri Ri Xin Publisher Ltd.
- Yip, Po-Ching, & Don, Rimmington (2004). *Chinese: A Comprehensive Grammar*. London & New York: Routledge.
- Zeitoun, Elizabeth (1992). *A Syntactic and Semantic Study of Tsou Focus System*. MA Dissertation. Hsinchu: National Tsing Hua University
- Zheng, Lin (1997). Tonal Aspects of Code-switching. *Monash University Linguistic Papers* 1:1, pp. 53-63.

## Appendix 1 – Letter of Consent

Dear Sir/Madam:

My name is Wang Sung Lan, a PhD student of the University of Wales, Bangor, UK. I am currently writing my PhD thesis and will probably need your assistance for collecting data.

The overall aim of my thesis is to investigate the current situation of language use in Taiwan. In order to gain the appropriate data for analysis, my assistants and I will be recording your conversation with your friends, family members or relatives at any possible informal occasions. The entire recording work will last 3-6 months in total. Your recorded speech will be transcribed into written document and digitalized as well as archived in electronic forms for possible later use on further language research.

If you agree to participate, please sign this letter and fill the questionnaire attached to this letter. Agreeing to participate in this research does not commit you to anything, and you may change your mind and withdraw at any time. Note that any reporting will be completely anonymous, and neither your names nor your personal details will feature in any reporting of this research.

We will be happy to discuss any aspect of the research with you, so if you have any questions or comments, please don't hesitate to inform us.

With best wishes,

Yours sincerely,

Wang Sung Lan

---

I understand that my speech will be recorded, and my name will not be revealed in any reports. I also agree that my recorded speech may be used later for archiving and for further language studies.

Signed:

Date:

## Appendix 2 - 語言學研究計畫參與同意書

您好：

本人王嵩嵐，目前就讀於英國威爾斯大學班哥分校之語言學博士班。由於我最近著手撰寫博士論文，所以必須從台灣蒐集資料。在蒐集資料的過程中，極需各位的協助。以下為我博士論文研究的主題、及研究過程之簡介：

此論文主要是探討目前台灣各種語言的使用現況，為了蒐集研究所需的資料，我們必須將您和親友，在任何非正式場合的談話內容，作全程的錄音。整個研究的錄音過程將長達 3-6 個月，而所有收錄的內容，將被轉換成書面及電子形式的文件，並保存於各國際語言學資料庫中，以提供未來任何語言學者作相關研究。

若您同意參與此研究，請在此同意書末端處簽名，同時協助我們填寫附件之問卷。若您在簽下此同意書後改變心意，您可於任何時間退出此研究。您的姓名及個人基本資料，將完全不對外公開。研究中任何有關您的敘述，都將以匿名的方式進行。若您對此研究有任何之批評與指教，歡迎隨時與我聯絡！

再次感謝您的幫助!!!

王嵩嵐 上

-----

本人 \_\_\_\_\_ (請簽名)同意我和親友間的談話內容被全程錄音，而我也瞭解我的名字及個人資料在研究中將不被公開，任何相關的討論都完全以匿名方式進行。同時，我也同意將所有錄下的談話內容，轉換成書面及電子型的文件，並保存於各國際語言學資料庫中，以提供各語言學者作相關研究。

請簽名：

日期：

### Appendix 3 – Questionnaire

Please provide the following information:

Your Name: \_\_\_\_\_

Male \_\_\_\_ Female \_\_\_\_

Occupation \_\_\_\_\_

Educational Background \_\_\_\_\_

Your Date of Birth (day/month/year): \_\_\_\_\_

Your place of Birth: \_\_\_\_\_

Present Address: \_\_\_\_\_

Your telephone number(s): \_\_\_\_\_

Your e-mail address: \_\_\_\_\_

1. What language do you usually speak at home? (Please tick your answers! You can tick more than one option.)

(a) Mandarin (b) Taiwanese (c) Tsou (d) Others (please specify) \_\_\_\_\_

2. If you are a student, please answer this question. If not, please go to the question 3.  
What language do you usually use when talking to your classmates or friends?

(a) Mandarin (b) Taiwanese (c) Tsou (d) Others (please specify) \_\_\_\_\_

3. What language do you usually use when talking to your friends?

(a) Mandarin (b) Taiwanese (c) Tsou (d) Others (please specify) \_\_\_\_\_

4. What language do you usually use in working place?

(a) Mandarin (b) Taiwanese (c) Tsou (d) Others (please specify) \_\_\_\_\_

(If you are a student, the working place then refers to the place you do the part time



job. If you don't have a part time job, please ignore this question.)

Thank you very much for your time.

---

## Appendix 4 – 各語言使用情形之問卷調查

請提供下列資訊：

姓名： \_\_\_\_\_

性別：男 \_\_\_ 女 \_\_\_

職業： \_\_\_\_\_

教育程度： \_\_\_\_\_

出生日期： \_\_\_\_\_

出生地： \_\_\_\_\_

住址： \_\_\_\_\_

電話： \_\_\_\_\_

E-mail: \_\_\_\_\_

1. 您在家都說什麼語言? (請勾選您的答案! 您可選擇兩個以上的答案)

(a) 國語 (b) 台語 (c) 鄒語 (d) 其他 (請說明) \_\_\_\_\_

2. 若您目前仍是學生，請回答此題，若否請直接回答第三題：

您通常和同學交談時，都使用何種語言?

(a) 國語 (b) 台語 (c) 鄒語 (d) 其他 (請說明) \_\_\_\_\_

3. 您通常和朋友(同學以外)交談時都使用何種語言?

(a) 國語 (b) 台語 (c) 鄒語 (d) 其他 (請說明) \_\_\_\_\_

4. 您通常在工作場合，都使用何種語言?

(a) 國語 (b) 台語 (c) 鄒語 (d) 其他 (請說明) \_\_\_\_\_

(若您目前仍為學生，則『工作場合』及指您打工的場所。若您沒打工則不用回

答此題)

非常感謝您的幫忙!!!

---

## Appendix 5: Bilingual Mandarin/Tsou clauses collected from Group 1

(normal print = Tsou; *Italic* = Mandarin)

---

1. o-suko                      tsʔo    *gan-jue-dao ta zai na-bian.*  
tense5, Non-Agent-2s    only    feel                      he at    there

‘You felt he (was) there.’

2. homo maitsa    *ren*    *yi-jing*    *zo*    *le.*  
if    like this    people    already    leave    perfective

‘If it is like this, then he has already left.’

3. micu    aumane                      *jiao ta de lao-po kai ban-yun-che ne!*  
then    after a while    ask    he poss Wife    drive    truck                      part-decl

‘Then after a while, (he) asked his wife to drive the truck.’

4. mitsueno *tong*    na    tseonu ?  
OK    open    Nom5 road

‘Is the road opened?’

5. moso                      *gao-zhuang* do    *lao-ban* ?o    X.  
tense1, Agent    tell tales    Obl2    Boss    Nom4 Name

‘Basuya told tales to his boss.’

6. moso                      *lian-dai-guan-xi* ?o    X & Y.  
tense1, Agent    relationship                      Nom4    Names.

‘X & Y were related to this event.’

7. *yi-lu-shang*    do    *yo*    *na-ge*                      *dao-zi*    emo    no    ie.  
over the road    all    have    that-class.    knife    knife    poss    House

‘There were knives and sheaths all over the road.’

8. *i-si-tsu eno sheng-dong-ji-xi do nia X.*  
 realis-3s-perf then sound-east-attack-west Obl2 dead name.

‘X used to use a kind of fighting strategy – ‘shen-don-ji-xi’.’

9. *mo hen pei-he ne mio X de*  
 tense2,agent very cooperative Nom7 at that time name poss

*ba-ba gen na-ge Y.*  
 Father and that-class. name

‘At that time, X’s father and Y were very cooperative.’

10. *feəʔə na-ge shi-ho hai yao na shui-jin.*  
 Pig that-class. time also should take tax

‘You should pay tax if you want to raise pigs at that time.’

11. *Keiaga zhu mo io?huge*  
 Hokka pig tense2,Aent concave area

‘Hokka’s pig was thin.’

12. *na-ge lian ho ci dodu zo-zo-de.*  
 That-class. face thick gen. Wrinkled wrinkled

‘The pig’s face is fat and wrinkled.’

13. *i-si tsumo yi-shi a!*  
 tense6, 3s have lost part., suggestion

‘He lost (something)!’

14. *magci moso no shang-diao ne?*  
 why tense1, Agent Obl3 hang part-Ques

‘Why did (he) hang himself?’

15. moso            *to-ting*    ho    i-si            *wen*            do  
 tense1,Agent eavesdrop and    tense6,3s interrogate    Obl2

*jing-cha*            do                    *fen-zhu-suo.*  
 police            Obl2                Police station

‘He eavesdropped the content of the policemen’s interrogation.’

16. hao-udo            *ge-bi*            ho    *to-ting*            ho    i-si  
 Go secretly    next door    and    eavesdrop    and    tense6,3s

*shen-wen*            do                    *keisatsu.*  
 interrogate    Obl2                Police

‘(He) went to the next door secretly and eavesdropped the police’s interrogation.’

17. X, Y    *hai-yo* Z            *de*    *ba-ba* *tsai* *ai-yu-zi*            *ione*    Shi-san-xi.  
 Name and    Name    poss    father    pick    plant name    in    place name

‘X, Y and Z’s father picked ai-yu-zi in Shi-san-xi.’

18. moh-cu                    *kai-shi*    ?o    X ho    Y    *kai-che*            *a!*  
 tense2, Agent-perf    start    Nom4 X and Y    drive-car    part.,RF

‘X and Y started to drive the car.’

19. *qi*    *mo-tuo-che*    ?o    X.  
 Ride motorbike    nom.    Name

‘X rode the motorbike.’

20. i-he                            *ai?di*    ?o            *os?o*    *suo-you*            *de*  
 tense6, NonAgent-3p see    Nom4    my    all            gen.

*dong-zuo.*  
 movement.

‘They saw what I did.’

21. *jie-guo ta-men zhang-hao l?əa do gong-che-zhan*  
then they just meet Obl2 bus stop

*xia-mian you yi-ge qiao.*  
Below have one-class bridge.

‘Then they just met at the bridge below the bus stop.’

22. X, Y, Z, ho nia W *hai-you* K ho io mamioi  
Name and dead Name and name and that old/senior

ci J de *ba-ba chuan-bu do yao zo-lu a!*  
gen Name poss father all all should walk part.,decl.

‘X, Y, Z and W (who passed away) and K and J’s father all should walk.’

23. *səmoef?hini do gong-lu.*  
follow Obl2 highway.

‘Follow the highway.’

24. i-o lua do ic?o *san-bu che-zi a!*  
tense6, Non-Agent-1s meet Obl2 at that time three-class. car part,RF

‘At that time, I met three cars.’

25. *manitsu moso-la melə sheng-huo ne*  
why tense1, Agent-hab. can survive part, Ques.

‘Why could we survive?’

26. aio *ba-ba do X moso-la di-yi-ge no*  
It is father Obl2 name tense1, Agent-hab first-class Ob3.

səmuħunu tanʔe.

distribute work here.

‘It was X’s father who firstly distributed the works to the people here.’

27. moh-la ahoi iuipasa ʔo X ta-men.

tense1, Agent-hab originally rich Nom4 Name they

‘X and other people were rich in the past.’

28. ie eno zo bu-shi kao zi-ji qu zao de shi da-jia  
This house cop neg-cop depend self go build nom cop everyone

*lai bang-mang.*

come help

‘Because of everyone’s help, the house could be built.’

29. you-shi-ho yi-ge-yue mei-you yi-mao-qian tetsʔo melə  
sometimes one month neg.-have one-hair-money still can

*sheng-huo.*

survive

‘Sometimes even if we didn’t have any money for a whole month, we still could survive.’

30. yi-ge-yue lun yi-ci do kyotokai.

One month turn once Obl2 cooperative society

‘Every month (different people) took turns to hold the cooperative club once.’

31. moh-la ionto bigi do kuiai ʔo nia mama u  
tense2, Agent-hab sit door Obl2 train Nom4 dead mother my



ho    zai    na-bian    ho    u?egətə.

and    at    there    and    sleep

‘My mother used to sit beside the door on the train and slept.’

32. *bu-hui    shuo    lea    iupa    totounu    de    xin-tai.*  
neg-can    say    like    each other    help    nom    attitude

‘It is not like that (they) would have the attitude of helping each other.’

33. *iola    hāno    ano    simeo    yi-tian    ne.*  
that    salary    cop    pork    one-day part-expl

‘The salary of one day was one chunk of pork.’

34. *ho    mi-cu    kyotokai    zou    ni-zi-ji    dai    bian-an.*  
If    tense4,Agent-perf    cooperative club    cop    yourself    bring    lunch box

‘If you were already a member of the cooperative club, you would bring your own lunch box.’

35. *lano    fang-zai    yi-qi    ?o    okazu.*  
cop    put    together    Nom4    dishes

‘Several dishes were put together.’

36. *a?nosi    ?o    bento    u    shi    liang-ge    dan    a!*  
usually    Nom4    lunch box    my    cop    two-class.    egg    part, inform

‘Usually, I bring two eggs in my lunch box.’

37. *maŋtʃi    mo-su-cu    laso    bonə    no    he-bao-dan?*  
Why    tense2, Agent-2s-perf.    can    eat    Obl3    dropped egg

‘Why could you eat dropped eggs?’

38. ?o fəzuiu da duo le.  
Nom4 egg big more part.

‘The eggs are much bigger.’

39. wa ma-ma hen hui aətʔətsə do sənoesənoecavə. mo  
My mother very good at raise Obl2 cattle tense2, Agent

hen hui yang  
very good at raise.

‘My mother is very good at raising cattle. (She is) very good at raising (cattle).’

40. ta hen hui yang ?o nia ino.  
She very good at raise Nom4 dead mother.

‘My mother (who was already dead) was good at raising (cattle).’

41. hoŋci aəmʔətə xiang-yao chi rou de shi-ho,  
If really want eat meat nom time

la-si haʔva ieeteuna ?o teoua ho ahilu.  
hab-2s gently gather Nom4 chickens and ducks

‘If we really wanted to eat meat, (my father) would gently gather the chickens and ducks.’

42. mo que-shao mo-lian ?e iuʔfafoinana maitanʔe  
tense2, Agent lack of training Nom1 young people nowadays

‘Nowadays, young people are lacking of training.’

43. *na-ge*      *io*   *tseonə*   *do shi*   *na-ge*   *xue*   *a.*  
 that-class.   that   road   all cop   that-class   snow   part, inform

‘In our childhood, the road was all (covered) by snow.’

44. *la*   *you*   *sheng-yin*   *homo*      *zo-lu,*   *you-mei-you?*  
 hab   Have   sound   if              walk   has-not-has

‘When we walked, there was some sound, wasn’t it?’

45. *psav?ohi*   *?o*      *shuang.*  
 break          Nom4   frost

‘The frost broke’.

46. *la*   *hen*   *piao-liang*   *io*      *bətənə.*   *i-sino*                      *tsfua*   *do*  
 hab.   very   beautiful   that   flower   tense6,Non-agent   cover   Obl2

*bing*   *a*  
 ice   part-inform.

‘The flowers were beautiful and covered by the ice.’

47. *zhe-ge*      *?e*      *di-qio,*   *yue-liang,*   *gen*   *tai-yang*   *mi-cu*  
 This-class   Nom1   earth   moon   and   sun   tense4, Agent-perf.

*jin*      *ju-li.*  
 close   distance

‘The distance between the earth, the moon and the sun has become closer.’

48. *wo*   *jue-de*   *mi-cu*                      *akei*   *maisa*   *de*   *iuovei*   *do*  
 I   feel   tense4, Agent-perf.   a little   seem   nom   back   Obl2

*nia*      *noanao*   *?e*      *ie*      *qi-wen.*  
 before   before   Nom1   sun   temperature

‘I feel the temperature of the sun seems to change back a little.’

49. *qi-shi* i-o *gan-jue-dao* ?e *ren* *maitan?*e *mio*  
 in fact tense6, Non-agent-1s feel Nom1 people nowadays already

*akei* *mioci* *iuovei* *ne* *noanao* *mio* *akei* *masate* *iuovei*.  
 a little want back Nom7 before already a little seem back

‘In fact, I feel nowadays people seem to want to go back to the past.’

50. *mo* *nana* *ɕunaŋuko* *namo* *guang-to* *da* *Tfuya*  
 tense2, Agent have ten something have without hair Obl1 place name

‘In Tfuya, ten something people cut all their hairs.’

51. *mo* *lio-xing* *do* *li* *guang-to* *da* *Tfuya*.  
 tense2, Agent popular Obl2 cut no hair Obl2 place name.

‘It was popular to cut off all the hairs in Tfuya.’

52. *la-he* *no* *na?na* *pua* *tang-ping*.  
 Hab-3p Obl3 often let lie down.

‘They usually let (him) lie down.’

53. *tena* *yuan* ?e *fuŋu*.  
 Will round Nom1 head.

‘(His) head will become rounded.’

54. *la-he* *iaŋtsa* ?o *wai-guo-ren* *ta-men* *bu* *xi-huan*  
 hab-3p say Nom4 foreigner they. neg like

*tang* *zhe* *shui*.  
 lie down asp,dur. sleep.

‘They say foreigners do not like to lie down when they sleep.’

55. ho pa *bian-bian-de* ?e fu<sup>2</sup>gu.  
If let flat-flat-nom Nom1 head

‘Let (his) head become flat.’

56. *jio-shi* panono ah?tu pəco-si!  
that was has neg. bottom-3s

‘(What I saw was) he did not have bottom.’

57. ts?o ahoi da *ho-mian* ho *ping* do *da-tui*.  
only from Obl1 back and flat obl. thigh

‘(The man’s) thigh and bottom was flat.’

58. mo ob?tsi ian?mum?a sapəci *nu-hai-zi*  
tense2,Agent half many hairs face girl

‘Half of the girl’s face was full of hairs.’

59. mo mas?a məu?məu no *ho-zi* ?e *yi-ban-si*  
tense2,Agent like hair Obl3 monkey Nom1 half-3s

‘Half of her face was full of hairs, just like monkey.’

60. dan-shi mo mev?tso<sup>2</sup>gu do ic?o mo  
but tense2, Agent marry Obl2 one tense2, Agent

*hen* *you-qian*.  
very rich

‘But the one she got married with was very rich.’

61. *fong* mo *wang* *xia*.  
wind tense2, Agent toward below

‘The wind blew toward below.’

62. homo bai-tian wen bu-dao zhu-zhe.  
If day time smell neg-arrive piggery

‘If in day time, you can’t smell the piggery.’

63. mo cong zhe-bian chui shang-chu.  
tense2, Agent from here blow upward

‘The wind blew from here and went upward.’

64. xiang wo jio wen-de-dao fuzu wang na-li pao.  
Like me then smell-can wild boar to where run

‘Like me, I can smell the wild boar, and know where it goes.’

65. mo gan-kuai ma-shang pao-diao a ?o X.  
tense2, Agent quickly right away run-away part-expl Nom4 Name

‘X ran away quickly.’

66. mo liang-ceng madae ?o i-si ioni.  
tense2, Agent two-level like this Nom4 tense6, Non-Agent-3s shelter

‘The shelter where the pig hid was between two layers of grass.’

67. ?e sikielia la kan ji-jie-xing la.  
Nom1 Plant name hab See season part.-affirm.

‘We can only pick sikielia in a certain season.’

68. gogoeno na-ge X cai ai-yu-zi cai le san-si-shi-nian.  
Therefore that-class. Name pick plant name pick perf. Three-forty-year

‘Therefore, that X has picked ai-yu-zi for thirty or forty years.’

69. mi-?o-cu ts?o dāmaevove do X, yiao-bu-ran  
tense4, Agent-1s-perf only return Obl2 place name otherwise

de-o eġətsə alə moġsi da fueġu.  
future-1s lost and cry Obl1 mountain.

‘I could only return to X, otherwise I would get lost in the mountain and cry.’

70. moġsi *shi guo-du jin-zhang a.*  
cry cop Over nervous part-com

‘Crying is because you are too nervous.’

72. *ni moc?o tai jin-zhang.*  
you just too nervous

‘You were just too nervous.’

73. *yo-yi-ci wo gen X ahoi boebono.*  
Once I and name start set up a trap

‘Once X and I started to set up a trap.’

74. *mi-cu eno elə a?o guo yi-zhen-zi.*  
ense4, Agent-perf. then smell I after a while

‘After a while, (the wild boar) smelled me.’

75. *ho mo elə da wei-dao da ren?*  
whether tense, Agent smell Obl1 smell Obl1 people

‘Wheter did (the wild boar) smell the people?’

76. *aomane yò yǒ yi-ci shan-zhu.*  
After a while again have once wild boar

‘After a while, the other wild boar came.’

77. *na-ge zhu-zi iho evi ta duan sheng-yin bi-jiao da.*  
that-class. bamboo and tree step break sound comparatively big

‘The wild boar made some noises by breaking the bamboo and trees. That noise was much louder.’

78. *de-ko no mainenu hoci ia cəmoi a?*  
will-2s Obl2 why if have bear part-Ques.

‘If you meet a bear, what will you do?’

79. *cəmoi hai bu-shi yi-yang hui wen-dao ren de*  
bear also neg.-cop. same will smell people poss

*wei-dao hui pao a.*  
smell will run part.-expl.

‘When the bear smells people, they will also run away!’

80. *mi-cu di-er-tian wo de shen-shang zhi you*  
tense4, Agent-perf. second day I poss body-above only have

*yi-ba dao*  
one-class knife

‘The second day came.’

81. *səfəŋnə shang-qu.*  
early morning up-go

‘I went up (to the mountain) in the early morning.’

82. *bosifo do leng-xian.*  
walk Obl2 chine

‘I walked to the chine, and saw that shelter.’

83. *mi-o-cu aobɔŋo ?o bonə do xi-fan.*  
tense4, Agent-1s-perf. quickly Nom4 eat Obl2 rice soup.

‘I ate the rice soup quickly.’



84. diapado gəbu-si ?o zhuang-bei-si ho matsutsumas.  
put in backpack-3s Nom4 equipment-3s and all the things.

‘I put his equipment and all the things into his backpack.’

85. mi-cu miumu danee xiao-pao-bu ne!  
tense4, Agent-perf face to here trot inform

‘I trotted toward here.’

86. mi-o-cu səcəhə do leng-xian.  
tense4, Agent-1s-perf. reach Obl2 chine

‘I arrived the chine.’

87. mi-o-cu səcəhə do ni zhu-she due-mian  
tense4, Agent-1s-perf. reach Obl2 your piggery opposite-side

a!

part-inform.

‘I went to the opposite side of your piggery.’

88. mi-o-cu zai na-bian ting-xia-lai kan.  
tense4, Agent-1s-perf at that-class stop see

‘I stopped there and saw (people).’

89. mi-o-cu səcəhə do X de liao-a a!  
tense4, Agent-1s-perf reach Obl2 Name poss shelter part.-inform.

‘I arrived X’s shelter.’

90. moso-la na na-ge səbəcə do zhang-nao.  
Tense1, Agent-hab take that-class. juice Obl2 camphor

‘(They) used to take the juice of camphor trees.’

91. moh-cu                    tsʔo   kao        do    *lin-ban-di*  
tense2, Agent-perf   only   depend on   Obl2   forest

ʔo            ceyâga.  
Nom4        Hakka

‘The Hakka people earned their livings depended on the forestry resources.’

92. *yi-qian da-bu-fen do shi*   tagpu   ʔo    *ling-yong-qian*.  
before   mostly   all cop   labor   Nom4   pocket money

‘Before most people earned some pocket money by doing some labor works.’

## Appendix 6: Bilingual Mandarin/Tsou clauses collected from Group 2

(normal print = Tsou; *Italic* = Mandarin)

---

1. *nae taiwan zhi-you wo-men* tsou emo aemətə ta chu-qu  
this Taiwan only we Tsou have really step out

*zhi yi bu.*  
this one step

‘In Taiwan, only we, Tsou people have really done something.’

2. ?o *da-wu-zu* momitsi *nong* do *zi-zhi-qu.*  
Nom4 Name want make Obl. autonomous-zone

‘The Da-wu tribe wants to establish an autonomous zone.’

3. aemətə *zhe-yi-ci* *wo-men* *yao* ta *chu-qu* zo.  
really this time we should step out walk

‘This time we really should do something for Tsou.’

4. *la-uso* *xi-wang* *wo-men* *zhe* *yi-bei* *de* *nian-qing-ren*  
perf-I hope we this one-generation nom young people

*do* *yao* *qu* *liao-jie.*  
all want go understand.

‘I hope the young people like our generation should understand it.’

5. *ta-men* *do* *hui* *qu* pa?tsohivi.  
they all will go church

‘They all go to the church.’

6. *dan-shi* *bu* *go* *shen-ru* da lenohi?ə.  
but neg enough deep Obl1 small clans

‘But it does not go deeply enough to small clans.’

7. *tsʔo Tfuya ho Tapangu ʔo zhu-dao.*  
only Name and Name Nom4 control

‘(The activity) is only controlled by Tfuya and Tapangu.’

8. *tsohivi da Tfuya micu aəmətə hen ji-ji.*  
know Obl1 Name already really very keen

‘I know Tfuya was very keen to control (the activity).’

9. *a-hoi do aʔkii-u ho amou-u zhe yi-dai o.*  
start Obl2 grandfather-my and father-my this one-generation part-inform

‘It started from my grandfather and my father’s generations.’

10. *cong wo jia-zu de guan-xi yi-zhi yan-shen*  
From my family poss. relationship continuously extend

*dao kuba.*  
to shrine

‘They are linked to my family and even to the shrine.’

11. *ta you xie mei you aəmətə zhi-xing.*  
3s have some neg have really operate

‘He didn’t complete some parts of the project.’

12. *bu-yao zai na-bian la etəʔiə tomo fæcəuiia tsi iudeu.*  
Neg-do at there pull pull that white poss clothe

‘Don’t go there and protest!’

13. *fong-guan-chu* i-he zhu-dao ?e  
Scenery management department tense6-3p lead Nom1

*zheng-ge zu-qun.*  
entire-class. group.

‘The scenery management department lead the entire group.’

14. *mas?a isits?o kong-zhi do fong-guan-chu.*  
seem only control Obl2 scenery management department

‘It seems (the whole place) is controlled by the scenery management department.’

15. *alano ta-men na-zhong ren li jio bai zai*  
cop they that-kind people benefit then put at

*qian-mian a!*  
front part.-justify.

‘That kind of people only care about their benefits.’

16. *alano bi-jiao gong-shi-hua.*  
cop comparatively stereo-type

‘It is more stereo type.’

17. *no aemətə siamate iəsə no iəsə tsou hote jiang-dao*  
Obl3 really should wear Obl3 clothes Tsou if speaking of

*zi-zhi.*  
autonomy

‘Speaking of autonomy, we really need to wear the traditional Tsou clothes.’

18. *er-qie hen zhu-zhong mamamoi.*  
also very respect old people

‘They also respect old people very much.’

19. *iane iaf?ana do shi guan-xin zo-zu de na-xie ren.*  
belong outsider all cop care Tsou nom that-class. people

‘They are all outsiders who care about Tsou.’

20. *hai yo yi-xie omoakei bobochio da Tsou de yi-xie ren.*  
Also have some a little know obl1 Tsou poss. some people

‘There are also some people who know a little about Tsou.’

21. *xian-zai maitane?e xin bu-shi zo zai tong yi-tiao lu.*  
now now heart neg-cop. walk at same one-class road

‘Now, their hearts are not together.’

22. *zai jia-shang iufafoinana hai-you xu-duo de jia-zu-zhang-lao.*  
Also add young people and many nom elders in the family

‘Also, young people and many elders in the family (can join the activity).’

23. *guang-shi zi-zhi-qu li-mian de budu jio hen*  
Only autonomous-zone inside ADM Chinese then very

*nan gao-ding le.*  
difficult deal with part.- complaint

‘The Chinese people in the autonomous zone are hard to deal with.’

24. *alə gen na-ge Xiao-pu de baba bao-yuan.*  
seriously to DM name poss. father complain.

‘(She) complained seriously to Xiao-pu’s father.’

25. *mioeno aələ dan zai tie-men ne.*  
really just block at iron-gate part-inform

‘My car just blocks the pass way.’

26. *iu?fafoinana de xiang-fa bu yi-yang a.*  
young people poss. thought neg. same part.-justify.

‘Young people’s thought is different.’

27. *atinghi iu?fafoinana ta yie bu xiao-de zi-ji de*  
But young people 3s also neg. know own poss

*na-ge zi-wo-yi-shi.*

DM. self-identity

‘But the young people also don’t have their own identity.’

28. *mainenu hao-xiang mamameoi gen wo-men wo-men de*  
Why seem old people and us us poss.

*zhong-jian hao-xiang duan-diao.*

between seem break

‘It seems there is a gap between old people and us.’

29. *yie bu shi maya.*  
also neg cop Japanese

‘It was not because of Japanese, either.’

**Appendix 7: Bilingual Mandarin/Southern Min Clauses Collected  
from Group 3**

(*Italic* =Mandarin; **Bold**= Southern Min; *Italic & Bold* = unable to identify)

---

1. **a**        **wa** **mu** **tzai**    **i** **e**    **bodı**    shi    shi-mo?

interj.    I    neg.    know    he poss.    Purpose Corp.    what

‘I don’t know what his purpose is.’

2. **da-gε**    **wu**    **da-gε**    **e**    *qu-yu*    **a!**

Everyone have    everyone    poss    Area    part.-affirm

‘Everyone has his/her own responsible area!’

3. **li**    **gin-la-ri**    **na gon** *mei you*    *chu-fa*        *mei you*    *fen*

you today    if say        neg    have        punishment neghave    divide

*qu-yu bo wei gon.*

area    neg.    word    say.

‘If there is no punishment and no personal responsible area being

divided, then I have no words to say.’

4. *er-qie*    *zhe-ge*    *dong-xi*    **hon**        *you jian*    *jio*    *you fa.*

also    this-class. thing    part-clarify    have award then have punishment.

‘Also, if you want to punish someone, you also need to give them some awards.’



5. *ying-gai ma e-dan a-nei-dʒɪ la!*

probably also can this part.-affirm.

‘Probably, things could also be done like this.’

6. *a-bo mi-giǎ dio-ai ga lan chu-fa.*

then thing should give people punishment

‘They give us nothing, but want to punish us.’

7. *li kuǎ in e hyon-dion ʃi an-na gu-li in e?*

You see they poss. Mayor corp. how encourage they how?

‘Do you see how their mayor encourage them (to do something)?’

8. *wan ai ti-yi zo zɛ lan gam-ga you*

We should suggest very many people feel have

*kun-nan la*

difficulty part-excl.

‘We should tell them that many people found it is difficult (to do this job)!’

9. *lon ga-bɛ li-wu bɛ san saŋ a.*

All take-buy gift buy what send part.-suggestion

‘Take all (the money to) buy them gifts or whatever.’

10. *ʃiam ʃu-lai gei-yu jiang-pin*

sign out-come give awards.

‘Sign (an official letter) and give (them) awards.’

11. **wa bo** *jiang-jin* **ma wu** *jiang-pin*

I neg. bonus still have awards

‘Although I don’t have any bonus, I still have some awards.’

12. **li hia** **ga-wu** **hi-le** *gong-wu-ren-yuan-qing-jia-tiao-li* ?

you there have or not DM staff of government agency-day off-regulation

‘Do you have the Staff of Government Agency’s Regulation of a day-off?’

13. *zui xin* **e** **ne** !

most new nom. Part.- decl.

‘(It’s) the latest one!’

14. **wa ze** **ji** **hi-le hi-le** *wang-lu shang zhua xia lai de*.

I this cop DM DM internet above catch down come nom.

‘This is what I download from the Internet.’

15. **wa** **d3ia** **lon** **ji** *qing-jia-fang-shi er-yi la* !

I here all cop. day off- method only part.-decl.

‘I only have the Methods for Requesting Day-off.’

16. *qi-shi xian-zai da-yu-jin* *bi-jiao* *shao* *yong-dao* **ne**!

In fact now bath towel comparatively less use part-excl

‘In fact, I don’t use bath towel very often!’

17. **wa ma lon bo in-dio yu-jin ne**

I also all neg. use bath towel part-excl

‘I don’t use bath towel at all, either.’

18. **lan mu-shi bo hi-le fu-li-hu-zhu a ni ?**

we neg.-cop neg DM welfare-cooperation part-excl part-Ques

‘We don’t have welfare cooperation (a kind of social welfare system), do we?’

19. **a hi gān wu lan di-gong nian-ji yue da ling**

DM that day have. people say age more big receive

*yue duo.*

more many

‘You know... on that day, some people say if you are older and you can receive more.’

20. **gin-la-li tjin ga bi-jiao xiu-xian.**

today wear part. comparatively casual.

‘Today, (you) wear more casually.’

21. **e! wa Ke Siji dʒɪ giā ji bɛ yong gua-hao gia**

Interj. my Name this class. cop should use express send

**a-shi bɛ?**

or should

‘Should I send Ke Siji’s case by express or...?’

22. **mu dio-ai hui lip-li hi-lə hi-lə zhang-hao lai-de.**

neg. should transfer in DM DM account in

‘Shouldn’t we transfer (the money) into (his) account?’

23. *lin-liao-lin liao-san-lin-jiu. de-tʃɪ dan liao-san-ling-liao. A mu zai-iā*

0-6-0 6-3-0-9 seventh year 6-3-0-6 DM neg know

**ʃi zo dʒɪ-go-wɛ a-ʃi lən-go-wei liā**

corp. treat one-class.-month or two months collect

‘060-6309. The seventh year (is) 6309. Well, I don’t know whether we should let him collect (the money) each month or every two months.’

24. *qian liao nian e mui dan a-nei? ... a wa a*

previous six year nom every year like this DM I really

**mu-zai an-na sən.**

neg.-know how count

‘Was it the same in the previous six years? I really don’t know how to count it.’

25. **i hi-lə hu-zhu-hwei e ri-qi shi qi-shi-yi nian.**

He DM Cooperative-Association gen. date cop. Seventy-one0 year

‘Well, the year he joined the cooperative association was 71 (1984).’

26. **a i e nian-ling lei?**

DM he gen. age part-Ques

‘Well, what about his age?’

27. *yi nian-ling lwe sən o?*

according to age down count part-ques.

‘We have to account (the amount of money he can collect) according to his age, don’t we?’

28. *i dio jio-yi nian giam si-zab-la.*

He then Ninety-one year minus four-ten-six.

‘Then, (we have to) count (the amount of money he can collect) by using 91 to deduct 46.’

29. *yong hi-lə ji-ben-shu lin-dian-ba-er ma!*

use DM basic counting number 0.82 part-expl.

‘You (have to) use the basic counting number 0.82.’

30. *a nian-ling yue da jio na-gə o?*

DM age more old then that-class. part-Ques

‘If you are older, can you get more?’

31. *jin-tian na yi-tai bu xiao-de due dʒɪ-diao*

today that one-Mclass8 neg. know which one-class.

**gim mu dio.**

nerve neg. right

‘(I) don’t know what’s wrong with that (printer).’

32. **jian-fi bo zhu-shə jing-pian i bo hua-do kuan.**

now neg. inject microchip he neg. way see

‘He can’t examine (the dog) because the microchip hasn’t been put into its body yet.’

33. **jing-pian fi zhu di he ba lai-dei e ne!**

microchip cop inject in DM meat inside nom part.-expl

‘You must detect (it by machine). Microchip is put into the body (of the dog).’

34. **i-he dʒɪ-de jing-pian zhu di he pi-fu-xia.**

DM one-class. microchip inject in DM skin-below

‘Well... one microchip is put under the skin.’

35. **you-shi-ho li nan du-a-ho i dan-an sun-gwei, li gə**

sometimes you if just he just finished inspection you then

**da-ki-li na jio mei-you-guan-xi**

post that then alright.

‘If you post (advertisement) right after he finished inspection, then that is OK.’

36. **wo gen ta shuo zui gao lio-qian, zui gao lio-chian**

I to he say most high six-thousand most high six-thousand

ho?

part-Ques.

ho?

'I told him (you can be fined) at most 6,000NT\$. 6,000NT\$ at most, right?'

37. **dan-dʒɪ-lei ran gio li chu ga ran goŋ mei-guan-xi.**

later people ask you pay. Give people say alright

'Later, the person may ask you to pay (the fine for him) because you told him it's alright.'

38. **zong-gong san wan... san si wan**

total three ten thousand three four ten thousand

**e ne.**

nom part.-excl.

'Totally, thirty thousand... thirty or forty thousand.'

39. **e wa ga li goŋ hi-gian ho hi-gian te dən-lai.**

DM I to you tell that-class. DM that-class. take back

**dui-dion be qing-song.**

captin will clean-send.

'Hey! Let me tell you ...take that (case) ...um...Take that back because our captain will re-arrange (it) and send (it to other department).'

40. **goŋ be due-zhi da xing le-se ga jia-ju la.**

Say will put big size rubbish and furniture part.-inform.

‘(They) say (they) will put big-sized rubbish and furniture (in an empty land).’

41. **yuan-qi goŋ be yong zi-yaun-hui-sho-chang**

Far-East Ltd. say will use resource recycling park

**hi-le kong-di**

that-class. empty-land

‘The staff of the Far-East Ltd. said (they) will use the empty land in the resource recycling park.’

42. **he gam ai sia hi-lə hi-lə shu-liang ?**

that whether should write DM DM quantity

‘Should (we) write the quantity of the (furniture)?’

43. **zi-yuan-hui-sho-chang mu ji di iŋ?**

resource recycling park neg. cop asp-dur use

‘Isn’t (someone) using the resource recycling park?’

44. **in bo di dui-zhi k-shio-hu e ne.**

They neg. dur place repairable nom. part-expl

‘They are not (offering the land) for putting the repairable furniture.’

45. **bin-iaŋ hi-de kong-di i goŋ e a.**



beside that-class. empty-land he/she say nom. Part.- softness.

‘It is (the land) beside that empty land, he says.’

46. *pang-bian bu shi an-zhao zhe-ci hi-lə he e kong-di ?*

beside neg cop according to this time DM that nom empty-land

‘(The land) beside that empty land is not (allocated) according to this plan.’

47. *zheng-li ga zheng-di shi bo gaŋ e.*

clean and arrange-land cop neg same nom.

‘To clean (the land) is not the same as to arrange (the land).’

48. *zheng-li shi bao-kuo tsao-a jia-mi loŋ ai kao-kao ki-lai.*

clean cop include grass everything all should dig-dig up.

‘To clean (the land) includes removing grass or digging everything up.’

49. *a zheng-di shi zuan-bo loŋ pɔo beŋ.*

DM. arrange-land cop. all all pave flat

‘Well... to arrange (the land) is to pave it flat.’

50. *bo di zheng-jian la.*

neg. dur construct part-expl.

‘No! (They are) not constructing (the land).’

51. *zheng-di ga zheng-jian tʃa ze a.*

arrange-land and construct differ largely part-expl

'To arrange (the land) differs largely from constructing (the land).'

52. **ki giə** *na-ge* **gaŋ-bu-zə** **ki ion e ne**

to ask that-class construction-department to use nom. part-expl

'(We) asked the construction department to construct (the land).'

53. **i gao** **dʒɪ-bin e** *luo-cha zə* **ze**

that come here nom drop very much

'The drop from there to here is very sharp.'

54. *er-qie zheng-di* **ho dio-ai** **ʃia zwi lau zui lo.**

and arrange-land DM should sluice water reserve water channel

'And arranging the land, (we) should reserve a water channel for the water to sluice.'

55. *zhe shi bi-ran* **e xian-xiang la.**

This cop. inevitable nom. phenomenon exclaim.

'This is an inevitable phenomenon.'

56. **li ga-mən ze zu-dʒi** **gui-han. wu san-er-lio bo san-san-lio.**

You ask this address which-lane. have 326 neg. 336.

'Ask (him) which lane he lives in. There is only lane 326 but no 336.'

57. **bo** *ni* *ka* *ga dui-zhang* **gɔŋ** **dʒia wan** **be** **ho** **wa** **kuan.**

neg you why to captain tell eat finish will let me see.

‘Then why did you tell the captain you will let me see it after eating (something)?’

58. **he** **di** *zhi zhi-qi-guan* **e.**

That for cure bronchitis nom.

‘That is for curing the bronchitis.’

59. **di** **gɔ** *ho-long* **e.**

for protect throat nom.

‘(That is) for protecting throat.’

60. *zhe-ge* *man hao-he de.* *yin-wei* *zhe-ge* **lai-de** **shi**

This-class. little tasty nom. Because this-class inside cop

*wu-mei han gan-cao de wei-dao.*

Black-plum and licorice nom flavor.

‘This (is) tasty because it has the flavor of black plum and licorice.’

61. **dʒin-dʒian** **e** *fu-qi* **o?**

really nom. couple part.- question

‘(Are they) really a couple?’

62. **a** **i** **dio-be** *jie-fu-sheng-zi,* *ran-ho ta-zi-ji* *yi-jing*

DM she prepare borrow-stomach-bear-baby then herself already

*huai-yun le.*

pregnant perf

‘Well, she planed to find someone to be a surrogate mother, but she was already pregnant.’

63. **ze due dʒɪ-le ji xiang-cai ?**

these which one-class. cop Coriander

‘Which one has coriander flavor?’

64. *xiang-cai ji dʒɪ-le.*

coriander cop this-class.

‘This is the one with coriander-flavor.’

65. *xiang-cai dio-ji en-sui e.*

coriander just-cop. coriander nom.

‘Coriander is coriander.’

67. **gi-ji wa ma gao bu qing-chu.**

In fact, I also sort neg clear

‘In fact, I am not sure, either.’

68. **i da-jie o?**

his eldest-sister part.-Ques.

‘(You mean) his eldest sister?’

69. **in nān-e ka-za ji jio qing-ren.**

They two-class before cop. old lover

‘They used to be a couple before.’

70. **dǐa shi-yi-dian ban nian ne.**

just eleven-o’clock half only part-compl.

‘(It’s) just eleven-thirty.’

71. **ni xian gei de ho?**

You first give nom. part-Ques.

‘(Did) you give (it to someone) first?’

72. **li... li mu ji be dai hui-qu e ni?**

You... you neg cop will take back nom. part-Ques.

‘Aren’t you taking it back?’

73. **ni yong yi-ban zua-dui bo gao la.**

You use half absolute neg. enough part-affirm.

‘Absolutely, (it is) not enough if you only use half of them.’

74. **le-to dio bo dia a. li ga-be tson jian?.**

lottery already neg win part.-compl You will do what

‘I didn’t get any money from the lottery. What do you want from me?’

75. a     **kuan wu hi-lə he zhen-kong e bo**  
DM. see cop DM. that vacuum nom. neg.

‘Well, see if there is any with vacuum packaging.’

76. i     **kwi-də liao ao jin-jin-chu-chu nən le-bai a ma.**

He surgery finish after in and out two week perfect affirm.

‘After the surgery, he was sent to the hospital for several times within two weeks.’

77. he     **wu shi-jian e la.**

that have time nom. part-impat

‘(It) requires some time (to recover).’

78. **dai-wan dʒin-ma zui chi-shiang e fi fa-mi kə?**

Taiwan now most popular nom. cop. What department.

‘Do you know what department is the most popular in Taiwan now?’

79. **gei dʒɪ-le sho-shen e.**

add one-class. plastic surgeon nom.

‘They have a new department...plastic surgeon.’

80. **dan-shi yie bu-nen ke-ze dʒia e i-fin la.**

However also neg-can blame these nom doctor part-com

‘However, we can’t blame these doctors.’

81. *yi-liao-shu-hu*    **fi**    **ai**    **gai-dzin**    *la.*

Medical-neglect    cop.    should    improve    part-compl.

‘Medical neglect should be avoided.’

82. **wa**    **gə**    *ju*    *ge*    *li*       **ho**    **li**    **tian**    *la.*

I    more    raise    class.    example    give    you    listen    part-expl.

‘Let me give you one more example.’

83. *bi-ru-shuo*    **li**    **dzi-le**    **lan**    *mo-ge*       *dong-xi*    **be**    **kui-da.**

for example    you    this-class.    Person    certain-class.    thing    will    surgery

‘For example, something wrong with your body, and then you’ll need a medical surgery.’

84. *ru-guo*    **di**    **bi-gə**       **i**    **e**    **gua**    *ta*    *hui.*

if    in    America    he    will    cut    he    will

‘In America, if you have such an illness, the doctors would decide to remove the tumor.’

85. **i**    *jing-zhui*    **sən**    **dzi-liap**    *zhong-liao.*

His    neck    grow    one-class    tumor.

‘(There is) one tumor on his neck.’

86. **i**    **ki**    **bi-gə**       *la.*       **dziə**    **fiao-in-bə.**       *yong*

He    go    America    part-inform    check    supersonic    diagnostic    set    use

*na-ge nei-shi-jing kuan.*

that-class. Endoscopies check.

‘He went to America. The doctors there used supersonic diagnostic set and endoscopies to check his body.’

87. *gao-bue-a yong nei-shi-jing ki-gua.*

to end use Endoscopies to cut

‘At the end, the doctor removed the tumor by using endoscopies.’

88. *gua-liao e t̃u-lai-liao. gao-fi hui-fu gə sɛn dʒi-liap*

cut-perf. take out-perf later recover again bear one-class.

‘(The tumor) were removed and taken out. But after he recovers, the tumor will appear again.’

92. *A i laŋ dio a-ne hue-fu zheng-chang.*

DM. he person then like this recover normal

‘Well... just like this, he recovered.’

93. *suo-yi ta-men zun-zhong i-jin e zhuan-ye pan-duan.*

so they respect doctor poss professional judgment.

‘So they respect doctors’ professional judgment.’



**Appendix 8: Bilingual Mandarin/Southern Min Clauses Collected  
from Group 4**

(*Italic* = Mandarin; **Bold** = Southern Min; *Italic & Bold* = unable to identify)

1. **in**    *er-zi*    **a-bue**            **bi-iab**            **e-kuan**            *o*  
his    son            not yet – neg.    graduate            seem            part- affirm.  
‘Her son has not graduated yet, hasn’t he?’
2. *qi-guai*,    *zhe-li*    *zen-mo*    *you*    *mai*    *hei-zhen-zhu* ? **gam-ji**            *jin-ko*  
strange    here    why            have    sell    fruit name ?    ques.-cop    import  
**e**            **a-ji** ?  
nom            or?  
‘It is strange. How come they sell he-zhen-zhu here? Are these fruits imported from other countries or?’
3. **in-gai**    *yi-nian-si-ji*                    **lon**    **wu**.  
should    one-year-four-season    all            have  
‘We have fruit in all seasons.’
4. *zheng-tong*    *de*            *yin-gai*    *shi*    *xia-tian*    *you*    **a**  
orthodox    nom    Should    cop    summer    have    part.-affirm.  
*dong-tian*    **ji**            **bo-in-gai**    **wu**.  
winter            cop    neg-should    have.  
‘Generally speaking, this kind of fruit only grows in summer, but not in winter.’

5. *na-ge jie-mu lai-de fang-wun ta shuo ho i-dzin lan lon*  
 that-class. program in interview him say interj before we all  
*zhi kai-fa na-ge xian-jiao, do mei-you kai-fa*  
 only develop that-class. banana but neg-have develop  
 man-guo  
 mango.

‘In that program, he was interviewed and said we only developed the techniques of growing banana but not mango in the past.’

6. *xian-zai ho zhi-dao taiwan you mang-guo, so-i ri-bun dio*  
 now interj know taiwan have mango, so Japan therefore  
*kai-fi dzɔp-ɟiu lan dai-wan e mang-guo ne.*  
 start receive our Taiwan poss. mango part.-excl.

‘Now, the Japanese knows Taiwan produces mango, so they start to import mangos from Taiwan.’

7. *dzin-dzian quan-nian shuei-guo dzia be-li ne!*  
 really whole-year fruit eat neg- finish part-excl

‘We can eat different kinds of fruit in the whole year.’

8. *dzin-ma gu-lai lai-ia la li-zi ho, a ge-lai*  
 now next peach part.-affirm peach part.-clarify DM next

**dio-fi** *na-ge* *da-hong-shi*.

is that-class. persimmon

‘The next one is peach, and then is persimmon.’

9. *xian-zai nai-ge tai-guo de ba-la* **dio bo-ki a.**

now that-class. Thailand poss. guava then neg.-go perf.

‘Now, Thailand’s bala has gone (no longer being sold).’

10. **lon** *gai-liang* **ki a.**

All modify-asp perf. part-affirm

‘(guavas) have all been (genetically) modified.’

11. **i-he** *pin-zhong* **bo-gaŋ e.**

That-class. species neg-same part.-affirm.

‘They are different of fruits.’

12. *na-ge jin-huang bu ke-yi wai-xiao* **la,** **in-wi hi-kuan**

that-class. Fruit Name neg. can export part.-affirm, because that-kind

**mi-giaŋ zo gin...**

object very fast

‘Jin-huang cannot be exported, because that kind of things will be (decayed) very fast.’

13. *hai han yuan* **ne.**

still very far part.-affirm

‘(The destination) is still very far.’

14. *lao-da*      **loŋ-bo**      **ʃu-ki**      *o?*

eldest son    all-neg    go-out    part.-question

‘(Her) oldest son has never gone out (studied abroad)?’

15. **a**      **in**      *di-di*      *di-xi-fu*      *ye do zai*

DM.    His    brother    brother’s wife    also also    part.-asp.-prog.

*jiao-shu.*

teach

‘Both his brother and his brother’s wife are teaching.’

16. **nən-e**      **loŋ-di**      **ga-tse, yo**      *bo-shi*      *shuo ni*      *you*

two-class. both-asp.-prog.    Teach and    neg.-be    say    you    have

*han da de*      **hi-lə.**

very    big    nom.    That-class.

‘Both of them are teachers, and they don’t have very big that (business).’

17. **mən-goŋ**      *fang-zi*      *chu-zu*      **gam**      **ʃi**      *zhao*      *gong-zheng-ren?*

ask-say    house    for-rent    if    cop.    find    referee

‘(They) asked if they needed to find a referee if they wanted to let the house.’

18. **hi**      **gaŋ di**      **hia**      **di**      *liao-tian.*

That    day    at    there    dur    chat

'We were chatting there on that day.'

19. *shi-ji-shang* X **goŋ be te, wa gam-ga zo gi-guai**

In fact Name say want retire I feel very strange

'In fact, I feel very strange that X said she wanted to retire.'

20. **wa kwan i dʒin-dʒian man wai-xiang-de.**

I see her really a little outgoing

'I think she is really outgoing.'

21. **i he zo huo-po zo wai-xiang.**

she that very active very outgoing

'She is really very active and outgoing.'

22. **yi-qian in ba-ba loŋ ti guang-to ma.**

Before her father always cut no hair part.- affirm.

'Her father always cut all his hair before.'

23. **in ba-ba loŋ kan dʒi-dai ka-da-tʃia, do zai**

Her father. always hold one-sclass8 bicycle always in

*yin-hang chu-ru.*

bank out-in

'His father always rode a bicycle and went to the bank.'

24. **in-wi** **dʒin-ma** **gɔ-lai** *ta-men* *liang-ge* *twe-le* **ma!**

Because now next they two-class. retire-perf part.-excl.

‘What comes next is those two will retire.’

25. X **mu** **ʃi** *ming-nian* *si-yue* **o?**

Name neg. cop. next year April part.-ques.

‘Isn’t X going to retire next April?’

26. **na** **mu** **ʃi** *shu-jia,* **dio-ʃi** *er-yue* **a.**

If neg. cop summer vacation, cop. February part.-affirm.

‘If not summer vacation, it will be February.’

27. *dong-tian* **ga** **wu** **on-lai?**

Winter whether have pineapple

‘We don’t have pineapple in winter, do we?’

28. *zhe-li* *you* *xi-gua* **nei.**

here have watermelon part-excl.

‘Here are some watermelons!’

29. *xi-gua* **dʒin-ma** **zo** **gui** **nei.**

watermelon now very expensive part.-excl

‘Watermelon is very expensive now!’

30. *wo zuo-tian xian mai na-ge, kan u dua-liap e bo.*

I yesterday want buy that-class. see have big-Sclass4 nom. neg

‘Yesterday, I wanted to buy that, so I checked if they had big one or not.’

31. *yao mai hi-lə fo-iə-a, kan yi-jin duo-shao.*

Want buy DM fruit name, see one-measure. how much

‘(I) wanted to buy fo-iə-a, so I asked how much it costs per 0.6 kilogram.’

32. *hi-lə gam ji gu-ni e shu-jia?*

That if. cop. last year nom summer vacation?

‘Is it last summer vacation?’

33. *hai mei you tai-fong yi-qian, na-ge shi-ho xi-gua bu-shi*

yet neg have typhoon before, that-class. moment watermelon neg.-cop.

*ye hen pian-yi, hə-dʒia gə din*

also very cheap tasty also sweet.

‘Before the typhoon came, the watermelons were very cheap, tasty and sweet, weren’t they?’

34. *be dʒi-de si-fen-zhi-yi cai san-si-shi kuai.*

Buy one-class. one fourth only three-four-ten dollar.

‘One fourth of a watermelon only costs thirty to forty NT dollars.’

35. *yi-jin er-shi-wu, kan qi-lai you be-be ho.*

One-measure. twenty-five, see up-come also white-white part.-compl

‘One watermelon is 25 NT dollars per 0.6 kilogram and it looks white.’

36. *zǒ hon-tai do pao-dao shui.*

Make typhoon all soak water

‘(watermelons) were all in the water when the typhoon came.’

37. *na a-nei on-ai xiu-xi. fǎ-bu-dē ai wu-fen-zhong ye ke-yi.*

If so have to rest about need five minutes also OK.

‘If you are tired, you have to take a rest. Even if you only take a rest for five minutes, it’s OK.’

38. *dao shi-ho jiao-liu-dao ai zuo-zhuan o.*

at moment intersection need turn left part.-remind.

‘At that time, you must turn left at the intersection.’

39. *in di jia-yo hio?*

they dur fill petroleum part-Ques.

‘The are filling the petroleum, aren’t they?’

40. *in-wi dǎ-gi sho-ji ho you dian gu-zhang, sǒ-i*

because this-class. mobile phone DM. have little problem so



**wa ga gon** *deng-xia hui shi-lian.*

I to say later will lose-contact.

‘There are some problems with this mobile phone, so I told her we will probably lose contact later.’

41. *ling-wai* **dʒɪ-dai** **ʃi** X *na-qu.*

The other one-class. Cop. Name take-go.

‘X took the other mobile phone.’

42. **a** *xin de dian-chi ye yao chong-dian-qi nei.*

DM new nom. battery also need charger part.-inform

‘New battery also needs charger.’

43. *kong-yun* **ʃa-bu-də** *san ge li-bai.*

Air-transportation about three class. Week.

‘Air mail takes about three weeks.’

44. *kong-yun* **dio-ʃi** *zhi-jie ga li gia gao li e di-dian.*

Air mail cop directly for you send to you gen. location.

‘Air mail means they will send the parcel directly to your place.’

45. **a li gam wu** *ding zui hao-hua e bo?*

DM. you if have book most luxury nom neg

‘Well, have you booked the most luxury room?’

46. **dʒin-ma** **gə** **di** *wai-huan-dao-lu* **nei,** **a-bue**  
now still in ring road part-inform not yet

*jin* *nan-er-gao*

enter south-second-super high way.

‘Now, we are not in the Second South Super High Way yet. We are in the ring road.’

47. *ge-ge* *di-fang* **lon** **wu** **a.**

various place all have part.-affirm.

‘Various places all have some tourist spots.’

48. *nao-zhong* **zɔ** **gin** **hai** **e** **nei!**

alarm clock very quickly break nom. part.-compl

‘The alarm clock breaks quickly.’

49. **dan-dʒɪ-le** **gon** *ni* *yao* *han-guo* *ko-wei* ?

wait-a while say you want Korean flavor

‘After a while, the waitress asked ‘do you want to try Korean food?’

50. **he** *ban-tiao* **wu** **dʒia** **tso.**

that noodle have like this thick

‘The noodle is as thick as this.’

51. **da-bai** **be** **dʒia** **dʒi-waŋ** **mi** **lon** *ai* *rao-chang.*

every time want eat one-class. noodle all have to drive around

‘Every time when we want eat noodles, we have to drive around (the city).’

52. **tʃa-bu-də ai**      *rao*                      *liang zho.*

About      have to      drive around      two      circles

‘We have to drive around the city about twice.’

53. **wan lon ki**      xin-guang-san-yue                      **dʒia**      *la-mian*      **nei.**

we      all      go      name of the department store      eat      noodle      part.-inform.

‘We usually go to Xing-guang-san-yue to eat noodles.’

54. **a li hi-lə diam-fin**      *na-ge dian-xin you she-mo?*

DM. You      DM. dessert      DM. dessert      have      what

‘Well... what kind of dessert do you have?’

55. *you ka-fei han jia-tang-bu-ding*      **a-ne**      **nia?**

Have      coffee      and      add-sugar-pudding      like this      only

‘Do you only have coffee and pudding?’

56. **wan xian-sheng ma lon di lao-din.**

my      husband      also      all      in      upstairs.

‘My husband was also in up-stairs.’

57. **i ion hi-lə** *su-jiao-tong*.

he use DM plastic vat

‘He used the plastic vat.’

58. **in-wi** *zhong-qio-jie* **be gao a**.

because mid-Autumn Festival will come part.-expl

‘(It’s) because the Mid-Autumn Festival is coming.’

59. *zhong-qio-jie* **wa nian-jia lon ban** *yan-huo*.

Mid-Autumn Festival my mother-home always put fire works

‘In Mid-autumn Festival, we usually play fire works in my parents’ home.’

60. **wa shi-yi dian shui**.

I eleven o’clock sleep

‘I go to bed at eleven o’clock.’

61. **te qi-shui lai lim**

take soft drink come drink

‘Have some soft drink.’

62. **i be bu-ding a?**

he want pudding part-Ques

‘Does he want some pudding?’

63. *ka-fei wa dio bo ai dǎia.*

coffee I just neg. like eat

'I don't like to drink coffee.'

64. *ta shuo yao shi-ban la.*

He say cost forty thousand part- inform

'He said it would cost forty thousand.'

65. X *gon hai mei jie-zhang.*

Name say yet neg. calculate money

'X said she hadn't calculated the money yet.'

66. *ji ti-yu-zu be ai la.*

cop athletic education department will want part.- agreement

'It was the athletic education department who said they wanted (it).'

67. *ni ru-guo zhe-yang wo jio dan da-a.*

you if like this I then overthrow desk

'If you do (something) like this, then I will break the desk.'

68. a *dǎin-ma xue-qi mo zai zuo zheng-ji.*

DM. Now semester end then do achievement

'Well, it is the end of the semester, so (she) wants to show some achievements.'

69. *yi ban kǎ gā hō i*

one ten thousand dollar then give him

'Then, (she) gave him ten thousand.'

70. *de-i dǔōn de-zui bu-qi e laŋ.*

first kind nettle neg.-up nomr person

'(He is) the first kind of person whom you can't nettle.'

71. *de-li dǔōn dio-fi nu-sheng.*

second class. cop. woman

'The second kind is a woman.'

72. *X be-daŋ tsam-ga xiao-wu-hui-yi ji bo?*

Name neg attend school-affair meeting cop neg.

'X can't attend the school-affair meeting, can she?'

73. *ka-za na wu jie-suan lon wu ho ti-yu-zu*

Before if have calculate all have give athletic education group

**ni?**

part-Ques.

'Does the school give enough budget to the athletic education department before?'

74. *ji an-na X di zā dǔō-fīā, a in fīā-mue di*

cop why name dur work chairperson DM. her sister dur

**zə**      *kuai-ji?*

work    cashier

‘How come X is the chairperson, and her sister is the cashier?’

75. *zhe-yang*   **gon**   **be**   **gue.**

Like this   say   neg.   pass.

‘This is not right.’

76. *er-qie*   **wa**   **ga**   **li**   **gon.**

also   I   to   you   tell

‘Also, let me tell you.’

77. **wa**   **ze**   **mu**   **zai-iaŋ**   *nei-qing*                      **e**   **laŋ.**

I   this   neg.   know   inside-information   nom   person

‘I am a person who doesn’t know the real situation.’

78. *ta*   *de*   **i-su**                      **gon**   **mu-men**                      **kai**   **hia-ze**   **jin**

he   poss   meaning   say   not necessary   spend   so much   money

‘What he means is that it is not necessary to spend so much money.’

79. **be**   **ho**   **in**   **ta**   *shuo-shi.*

will   let   them   study   master

‘(He) will let them to study master.’

80. **dʒɪ-ben ki pa** Wu-shan **hon.**

Once go climb Black Mountain part-inform

'One time, (we) climbed the Black Mountain together.'

81. **dʒiao-gon ze lon xin-tai nei.**

generally speaking this all the way he thinks part-comt

'Generally speaking, this is all because of the way he thinks.'

82. *cong xin-li-xue lai gon in-gai ji wu*

from psychology come say may cop. have

*sho-guo cuo-zhe.*

experience-exp failure

'From the perspective of psychology, (he) might experience some failures before.'

83. **hi-ben X xiao-jie in hao-sɛn Y gɛ-hun.**

that time Name Ms. her son Name get married

'Last time, Ms. X's son Y got married.'

84. **fjiao-gue shi-jian i dio ka-dian-we gu-lai.**

Over time she then call come

'If (her husband goes home a little late), she will call me.'



85. *ta jio da-dian-hua shuo lin dao shi-zhong paī-ki a*

She then call say your home clock broken part-affirm

**hon**

part-Ques

‘Then, she called me and said your clock was not working, was it?’

86. *a i dʒɪ duan ʃi-gan dio ga X zi-dong gon*

DM. he one period time then to name automatical say

**wa lən le-bai mai pa.**

I two week neg. play

‘After a period of time, he would then said to X that I would not play (mahjong)

for two weeks.’

87. *i bue-a wu ka fang-kuan a la!*

She finally have more loosen part-RF part-comt

‘Finally, she was less strict.’

88. *lin lao-gong bo lai o?*

your husband neg. come part.- ques.

‘Why didn’t your husband come?’

89. *i lon gon gian i e jio-bing gə fu-fa.*

He always say afraid she poss old-illness again relapse

‘He always says that he is afraid that her illness might relapse.’

90. **ka-za** **X** **gam** **ji** **e** **fia** *zhuo-zi* **a?**

before Name if cop. would throw table part-Ques

‘X would throw the table before, wouldn’t he?’

91. **i** **wu** **dʒɪ-le** *dong-zuo* **zɔ** *hə-ʈʃiə*.

he have one-class. movement very funny.

‘He had a very funny expression.’

92. **i** **kə-lin** *guan-nian lai-de* **lon** **gon** **wu** **gui** **di** **ga**

He probably notion inside all say have ghost dur give

**sa** *ma*.

pull part.-explain.

‘Probably, he thought that the ghost was pulling him.’

**Appendix 9: Bilingual Mandarin/Southern Min Clauses Collected  
from Group 5**

*(Italic =Mandarin; Bold= Southern Min; Italic & Bold = unable to identify)*

---

1. **in-wi**    **i**    *mai-ke-fong*    **din-i**    **di** **ɕʒia**    *la*            **gə**
- because    its    microphone    equal    at here    part-expl    and
- di**    **ɕʒia**    *la*.
- at    here    part-expl
- ‘It’s because its microphones are at here and here.’

2. **a**    *ni*    *shuo na*    *ge*    **lin**    **dao**    **ɕʒia**    **hə-liao**
- DM    you    say    that    class    your    house    eat    delicious food
- lon**    **an-na**    **ɕʒia**.
- all    how    eat
- ‘Well... You said you had delicious food with your (flat mates). What kind of food did you eat?’

3. **do**    *pin-ming*    **ɕʒia**    *a*.
- all    try hard    eat    part-expl
- ‘(We) always try very hard to eat (everything).’

4. *hao-xiang*    **kə**    **di-le**    **də**    **din**.
- Seem    hit    at    table above

‘It seemed I hit the table.’

5. *shang-ci wo-men jio yi tua ren qu chi a.*

last time we then one group people go eat part-excl.

‘Last time a whole group of us had a meal together in a restaurant.’

6. *ran-ho ke-neng wen ge-bi be ki bo?*

then possibly ask next door want go neg

‘Then, we will probably ask our neighbours if they want to go or not.’

7. *li mu shi be can-jia jiao-hui?*

You neg cop will join church

‘Aren’t you going to the church?’

8. *A mu shi be zu yang-rou-lu.*

DM. neg. cop will cook lamb-fondue.

‘Well, aren’t you going to cook lamb-fondue?’

9. *ai jin go-tong o ha la.*

should first communicate DM good part-RF.

‘You should communicate with them beforehand.’

10. *i dio zao-ki hi-le lan chuang wu bo?*

she then go to DM. Hit bed have neg.

'She then hit the bed, didn't she?'

11. *ran-ho ta jio bu shuang shuo ze mu-fi lin dao*

then she then neg happy say this neg-cop your house

**li di tse sā-jiao.**

you dur adjust what

'Then she wasn't very happy and say, 'This is not your house. How can you adjust (the air conditioning without my permission?)'

12. *ran-ho i dio gon i ni chi bao lo?*

then she then say DM you eat full part.-Ques.

Then she said, 'um?... Have you had enough?'

13. **i dio i-di pai-ji X**

she then all the time alienate name.

'She was alienated (by her flat mates) all the time.'

14. **hi-gin ji shi-zhang-shi dio bo.**

that-Sclass7 cop flat-leader-room yes neg.

'That is the flat leader's room, isn't it?'

15. *ru-gu be huan he qin-shi e ji an-nua.*

if want change that room nom cop what

'How to change the room?'

16. **li be ham i ki guo-ye o?**

you will with him go stay over night part-Ques.

‘Will you sleep with him?’

17. **i dio ji jio shi guo-jun-yi-yuan dui-mian you**

it then cop then cop National Army Hospital opposite have

*yi tiao xiao lu a.*

one Mclass6 small road part-expl

‘There is a small road on the opposite side of the National Army Hospital.’

18. **bao-ya dam dzia e-hua a!**

Shop name at here if part-

‘If Bao-ya is here...’

19. **bo lan ga wun guo-ye a!**

neg. person with us stay over night part-comp

‘No one wants to sleep with us!’

20. **li mu-fi gon min-ia-zai be gam i guo-yie ma?**

You neg.-cop. say tomorrow will with him stay over night part.-Ques.

‘Didn’t you say you will sleep with him tomorrow?’

21. **tsai-tji-a hen pian-yi.**

market very cheap

‘Things are very cheap in that market.’

22. **wa** *tong-xue* **gon** **e**.

my classmate say nom.

‘My classmate said so.’

23. **lin** *qi-yi-guo* **dzi-lap** **da-an** **go** **ko** **a-fi** **la** **ko**.

your kiwi fruit one-Sclass4 only five dollar or six dollar

‘One kiwi fruit only costs five or six dollar in the place you live’

24. *ying-xin* **na** **wu** **sā-mi** **ho-kuang?**

Welcome-new what have what good-looking

‘Why is the welcome party interesting?’

25. **X** *shuo* **ta** *kan-dao* **yi** **ge** **a-be** **a**.

Name say she see one class old man part-report

‘X said she saw an old man.’

26. *ran-ho* **jio** **du-gu** **a!**

then then doze part-expl

‘Then he dozed.’

27. *xiang* *bu-dao* **li** **ji** **dzi** **le** *xian-qi-liang-mu* **ne!**

think neg. you cop one class good-wife-good-mother part.-comt

‘I don’t know you are such a good wife and such a good mother.’

28. **li ji jun be zə dʒɪ dʒɪŋ la-ta e**

You cop want will be this kind sloppy nom

za-bo-lan ma?

woman part-Ques

‘Do you want to be this kind of sloppy woman?’

29. **zui du fu-ren xin li wo tian gue bo?**

most poisonous woman heart you have hear perf neg

‘Have you ever heard the idiom “women have the most poisonous heart?”’

30. **wa dʒin kai-fi ga li xun-lian ho-a la.**

I now start giver you train good part-sug

‘I’ll start to train you from now?’

31. **zhe ge ka wu jin li.**

This class more have like you

‘This is more like what you would say.’

32. **wa e jian-bang wu-gao sən e la.**

I poss. shoulder have sore nom part-compl

‘My shoulders are sore.’

33. **na ge ko-qi dʒin-dʒian zə zan e.**

that class. manner really very good nom



‘The way you spoken was really cool.’

34. *ban-niang qu le nei!*

half-year go perf part-surp

‘Half a year has passed!’

35. *da-gai mei ke ai ji fen ka e-dǎ lu-qu.*

About every subject require what score then can recruit

‘How many scores do you need to get for each subjects in order to be recruited?’

36. *ru-guo bo ping-jun dio be a-nua?*

If neg. average achieve will what

‘If your average score is still below the minimal requirement, what will you do?’

37. *hu-li shi ke san-mi?*

nursing cop. test what

‘What kind of questions will appear in your nursing test?’

38. *lin gam ai ta hua-xue?*

You if need study chemistry

‘Do you need to study chemistry?’

39. *shang zhuan-ke lon ka dǎo di bue.*

attend junior college all comparatively less dur memorize

‘I seldom memorize (the technical terms) after studying in the junior college.’

40. **lan** *ying-wen* **lon** **ji** **ai** **kau** **bue** **ba.**

our English all cop. should depend memorize part.-surmise

‘The only method to study English is memorising (vocabularies).’

41. *na* *hai* *bu-ru* **mai** **ta** **dio** **bo?**

then also rather neg. study right neg.

‘Then it’s better not to study it, right?’

42. **li** **ze** *xiang-lian* **ji** **di** **due** **be** **e?**

you this necklace cop at where buy nom.

‘Where did you buy this necklace?’

43. **dʒɪ** **dʒon** *liang-liang-de* **o?**

this kind shiny part-Ques.?

‘Do you mean this kind of shiny one?’

44. *ran-ho* **wan** *lao-shi* **dio** **gon** **wan** **lon** **zo** **sfin-ko** **e.**

then our teacher then say we all very poor nom

‘Then our teacher said you are very poor.’

45. **dio** **ʃun** **he** *bai-hua-wen* **wu** **bo?**

just like that modern Chinese have neg.

‘Just like modern Chinese, right?’

46. **he** *gu-wen-lei* **e** **mi-gian** *o?*

that classical Chinese nom thing part-Ques.

‘Just like classical Chinese?’

47. *wo jue-de chu-lai* **da-ge** **lon ion** *wen-xue-lei de.*

I feel go out everyone all use literature kind nom

‘I feel in the job market people like to recruit those who study literature.’

48. *ran-ho* **jia ge** **jia he** *gu-wen de.*

Then write also write DM old English nom

‘Then they all wrote in old English.’

49. *gao-shang de shi* **di** *xiu wen-xue e.*

elegant nom. cop. dur. study literature nom.

‘Those who study literature are elegant.’

50. **lin he** *o yi qun mao go.*

You kind DM one group cat dog

‘You are like a group of cats and dogs.’

51. **li ki be** *lu-yin-bi o?*

you go buy recording-pen part-Ques.

‘You bought a digital recorder?’

52. **a li ma-fi e-dang di dian-nao shang ting.**

DM you also can at computer up listen

‘Well, you can also use the computer to listen to your recordings.’

53. **in-wi wan dao dian-nao gua-diao.**

Because my home computer die

‘It’s because my computer is dead.’

54. **wa bun-lai ji bē bè ze lai lu wo-men shang-ke de.**

I originally cop. will buy this come record our lecture nom.

‘Originally, I bought this digital recorder to record our lectures.’

55. **wui tao-za gao-diam ta ga shi-er-dian.**

from morning nine o’clock study to twelve o’clock.

‘My lecture started from nine o’clock in the morning to twelve o’clock.’

56. **wa ma-ma dio di mən gon wa gam ai ki bo**

My mother just dur. ask say I if need go study

*shuo-shi e he.*

master poss. that.

‘My mother asked me if I needed to prepare the test for the master course.’

57. **li nan kao shuo-shi bu do you qu bu-xi ma?**

You if test master neg. all have go extension lesson part.-Ques.

‘Shouldn’t you go to the extension lessons for preparing the exam of the master course?’

58. **wa dio kuan-dio he xue-sheng a.**

I then see those students part-expl

‘Then I saw those students.’

59. **zao-shang gao e-bo, ran-ho wan-shang ge wan-zi-xi**

morning to afternoon then evening again late-self-study

**dao shi-diang nei!**

to eleven o’clock part-compl

‘They study from morning to the afternoon, and then do self-study until eleven o’clock in the evening.’

60. **wa gam-ga he ji dzi le qu-shi.**

I feel that cop. one kind trend

‘I think that is a kind of trend.’

61. **dziin-ma lan lon zo zhu-zhong e.**

now people all very take seriously nom.

‘Now people all take it very seriously.’

62. *jio-shi shuo li na ji be di da yi-yuan de-hua,*

just-cop. say you if cop want in big hospital if

*ta de xu-qio bu duo.*

he poss. demand neg much.

‘That’s to say the job vacancies in big hospitals are limited.’

63. *gu-ge de ying-wen nei-zang a gan-zhang a*

bones poss. English internal-organs part.-listing liver part.-listing

*jia-mi jia-mi a-ge wu dong-wu e lon wu e-dio.*

etc. etc. also have animal poss. all have learn

‘We have learned the English words for human beings’ bones, organs, liver, etc as well as those for animals’.

64. *wo jin-tian bei-yi-bei kao wan dio bue-gi a.*

I today memorize test finish then neg.-remember part-expl

‘Although I memorize a lot now, I will forget everything after the test.’

65. *ke-shi ru-guo bī-ian e lan lon a-ne e ji-zun,*

but if surrounding poss. people all like this poss moment

*i e cheng-wei quan-min-gong-di.*

he will become enemy

‘But if people around her are like this, she will become their enemy.’

66. *chu-fei shuo zheng dong lon ji lin e lan, li ka*

only if say whole class all cop you poss people you then

e-dan a-ne zə.

can like this do.

‘You can do something like this only if all people living in this building are your friends.’

67. **he ji wang-lu-fei.**

That cop internet-fee

‘That is for internet fees.’

68. *na ying-gai bao-gua zai dian-hua-fei li-mian,*

that should include in telephone bill inside

**li na ji wu den-wei de-hua.**

You if cop. have telephone if

‘That should be included in the telephone bill if you have a telephone.’

69. *ru-guo li na be ga i sha gao san-qian-wu-bai*

if you if want make He lower to three-thousand-five-hundred

**hə a, ran-ho han-shu-jia i bo ban-dio san.**

good part-RF then winter-summer-vacation he neg help anything

‘If you want him to lower the rent to 3,500NT\$, then he won’t help you anything in winter and summer vacation.’

70. *ru-guo zhu-xiao de-hua, ma shi jin e-hə a.*

if live-school if also cop really reasonable part-comt.

‘If you live in the dormitory, the price is also reasonable.’

71. *er-qie ta han-shu-jia bo ga li sən dʒin.*

Also he winter-summer-vacation neg. give you count money

‘Also, he does not ask you to pay the rent in winter and summer vacations.’

72. *li sən yi ge yue wa-ze?*

You count one class month how much

‘How much do you need to pay for one month?’

73. *lan dua si ge yue mu shi?*

we live four class month neg cop.

‘We will live there for four months, won’t we?’

74. *li gon hi-le si-qian-kuai e ma.*

you say that four-thousand-dollar nom. part-cla

‘You said that cost 4000 dollars, didn’t you?’

75. *bu-guo i wu fu leng-qi tao-fang ma.*

but he have include air conditioning room part-cla

‘But (his house) includes a room with an air conditioning.’



76. **dʒia-ni-hə!**      **gə wu chu-fang.**

how wonderful    also have kitchen.

‘How wonderful! (This house) also has a kitchen.’

77. **i gə wu he ting-che-chang.**

It also have DM parking space

‘It also has a parking space.’

78. **yun-dong yi-xia ma hə.**

exercise a-little also good

‘It is also good to have a little exercises.’

79. **wa be ion you-cun e mi-gĩa.**

I need use post-savings poss thing

‘I need to deal with my savings in the post office.’

80. **wa san-dian duo dio xia-ke a!**

I three-o’clock more then finish-lecture part-expl

‘I will finish my lectures around three o’clock.’

81. **wa yi xia-ke dio ka den-wei ho li.**

I one finish-lecture then hit telephone give you

‘I will give you a call right after finishing my lectures.’

82. *wo ke-neng qing X ki kuan.*

I probably please Name go see

'I will probably ask X to have a look.'

83. *wa ji jun gon han wu-zhu dua he tong yi*

I cop think say with landlord live DM same one

*men-hu dio bo he a*

door-window then neg. good part-comt

'I think it's not good to live with the landlord.'

84. *dian-mian ji de be jia e?*

store-face cop dur. sell what nom.

'What does that store in the first floor sell?'

85. *na ji be jia guai-guai e mi-gia.*

If cop sell what strange nom thing

'What if it sells some strange things?'

86. *bo lan jun be ga i er-qie bu yao zai ti-dao*

neg. person want will with her and neg will again mention

*ta le*

her part-req.

'No one wants to (live) with her, and don't mentioned her again.'

87.wa bo gon be ham i guo-yie.

I Neg. say will with him sleep

‘I didn’t say I would sleep with him.’