Pacific Linguistics E-1

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

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

Copyright in this edition is vested with Pacific Linguistics
First published 2005

National Library of Australia Cataloguing-in-Publication entry:


SEALS XV : papers from the meeting of the Southeast Asian Linguistics Society.

Bibliography.
ISBN 0 85883 563 0.


495
# Contents

Contents   iv
Program of the SEALSXV meeting   vi
Editor’s preface   ix

## Syntax and lexicon

Taboo object relative clauses in Indonesian  
*Tim Hassall*  
1

The invisible agent with global meaning: Thai zero anaphor subjects  
*Charatdao Intratat*  
19

Shape-based classifiers in Hakka and comparison among three dialects  
*Chiachun Lee*  
29

The logical constructions of conditional sentences and the functions of conditional expressions: a contrastive study of Japanese and Thai  
*Saowaree Nakagawa*  
41

Errors in the translation of topic-comment structures of Vietnamese into English  
*Pham Phu Quynh Na*  
51

Nominalization and categorization of verbs in Thai  
*Amara Prasithratsint*  
73

Negation in Kavalan  
*Li-May Sung and Yuting Yeh*  
83

Preliminary observations on a “migrant Vietnamese”  
*Bao Duy Thai*  
97

The allative preposition in Thai  
*Kiyoko Takahashi*  
111

Pronoun retention in Khmer and Thai relative clauses  
*Natchanan Yaowapat*  
121

## Phonology

Interaction of phonology and semantics of Taiwanese’ names  
– an optimality theory approach  
*Chi-hua Hsiao*  
133

The tone and syllable structure of Turung  
*Stephen Morey*  
149
Comparative Mon-Khmer

The Vieto-Katuic hypothesis: lexical evidence
Mark J. Alves 169

Tone correspondences and tonogenesis in Vietic
Koichi Honda 177

Proto-Katuic phonology and the sub-grouping of Mon-Khmer languages
Paul Sidwell 193
# Program of the SEALSXV meeting

**Wednesday 20 April 2005**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session 1: Coombs Lecture Theatre</th>
<th>Session 2: Seminar Room A</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.30-9.15</td>
<td>Registration: Coombs Lecture Theatre Foyer</td>
<td></td>
</tr>
<tr>
<td>9.15-9.30</td>
<td>Opening welcome</td>
<td></td>
</tr>
<tr>
<td>9.30-10.30</td>
<td>Plenary Session</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Justin Watkins, The SOAS Wa Dictionary Project: Lexicography for a Divided Language</td>
<td></td>
</tr>
<tr>
<td>10.30-11.00</td>
<td>Break: refreshments in Coombs Lecture Theatre Foyer</td>
<td></td>
</tr>
<tr>
<td>11.00-11.30</td>
<td>Mark Alves Vieto-Katuic and the lexical evidence (presented on video)</td>
<td>Jayyoung Choi and Lori Fredricks Integrating the Four Skills with Lower Level Learners: Techniques for Teaching ESOL and Korean with Beginning Students</td>
</tr>
<tr>
<td>12.00-12.30</td>
<td>Koichi Honda Tone Correspondences and Tonogenesis in the Vietic Family</td>
<td>Norhaida Aman The Acquisition of Colloquial Singapore Malay</td>
</tr>
<tr>
<td>12.30-13.00</td>
<td>Lunch: Caterina's Cafe</td>
<td></td>
</tr>
<tr>
<td>13.30-14.00</td>
<td>Ross Slater Kui and Thai — aspects of their relationship</td>
<td>Alice Yin Wa Chan Cantonese ESL Learners’ Acquisition of Tri-literal and Bi-literal English Onsets: An Empirical Study</td>
</tr>
<tr>
<td>14.00-14.30</td>
<td>Pascale Jacq The who/what and how of Nyaheun $hɨ$.</td>
<td>May Lee Do teachers of English in Hong Kong teach reading strategies they themselves use?</td>
</tr>
<tr>
<td>14.30-15.00</td>
<td>Natchanan Yaowapat Pronoun Retention in Khmer and Thai Relative Clauses</td>
<td>Heather Winskel Language acquisition in the temporal domain.</td>
</tr>
<tr>
<td>15.00-15.30</td>
<td>Break: refreshments in Coombs Lecture Theatre Foyer</td>
<td></td>
</tr>
<tr>
<td>15.30-16.00</td>
<td>Meng Vong The Criteria of Compound Words in the Modern Khmer Language</td>
<td>Bao Thai Preliminary observations on Vietnamese as a migrant language</td>
</tr>
<tr>
<td>16.00-16.30</td>
<td>Chirapat Prapandvidya Indic Origin of Some Thai Words In Early Inscriptions and Early Literature in Thai Language.</td>
<td>Li-May Sung and Yuting Yeh Negation in Kavalan (read by Chi-hua Hsiao)</td>
</tr>
<tr>
<td></td>
<td>Conference Dinner – at the Lemon Grass Restaurant</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Session 1: Coombs Lecture Theatre</td>
<td>Session 2: Seminar Room A</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>8.45-9.30</td>
<td>Registration: Coombs Lecture Theatre Foyer</td>
<td></td>
</tr>
<tr>
<td>9.30-10.30</td>
<td>Plenary Session</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graham Thurgood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>From Proto-Chamic to Tsat:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Insights from Zheng 1997 and from Summer 2004 fieldwork.</td>
<td></td>
</tr>
<tr>
<td>11.00-11.30</td>
<td>Pittayawat Pittayaporn</td>
<td>Charatdao Intratat</td>
</tr>
<tr>
<td></td>
<td>Moken as a Mainland SEA language</td>
<td>The Invisible Agent With Global Meaning: Thai Zero Anaphor Subjects</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td>Stephen Morey</td>
<td>Amara Prasithathsint</td>
</tr>
<tr>
<td></td>
<td>The Turung tonal system – when</td>
<td>Nominalization and categorization of verbs in Thai</td>
</tr>
<tr>
<td></td>
<td>tones lose functional load.</td>
<td></td>
</tr>
<tr>
<td>12.00-12.30</td>
<td>Zaharani Ahmad</td>
<td>Saowaree Nakagawa</td>
</tr>
<tr>
<td></td>
<td>Lunch: Caterina's Cafe</td>
<td></td>
</tr>
<tr>
<td>13.30-14.00</td>
<td>Emmie Li</td>
<td>Tim Hassall</td>
</tr>
<tr>
<td></td>
<td>A Constraint Based Analysis of Syllable Contraction: A Case Study on Mandarin Chinese</td>
<td>Taboo object relative clauses in Indonesian</td>
</tr>
<tr>
<td>14.00-14.30</td>
<td>Kimiko Tsukada</td>
<td>Rodney Jubilado</td>
</tr>
<tr>
<td></td>
<td>English and Thai vowels produced by Thai bilinguals - the effect of following stops</td>
<td>Bahasa Isamal: The Syntax of a Dying Language</td>
</tr>
<tr>
<td>14.30-15.00</td>
<td>Chi-hua Hsiao</td>
<td>Chiachun Lee</td>
</tr>
<tr>
<td></td>
<td>Interaction of Phonology and Semantics of Taiwaners’ Names - An Optimality Theory approach</td>
<td>Shape-based classifiers in Hakka and comparison among three dialects</td>
</tr>
<tr>
<td></td>
<td>Lunch: refreshments in Coombs Lecture Theatre Foyer</td>
<td></td>
</tr>
<tr>
<td>15.30-16.00</td>
<td>Ben Arps</td>
<td>Mark Post</td>
</tr>
<tr>
<td></td>
<td>Media in the making of a ‘regional language’ for Banyuwangi (East Java), c. 1970–2004</td>
<td>Compounding and the structure of the Tani lexicon</td>
</tr>
<tr>
<td>16.00-16.30</td>
<td>Natthaporn Panpoothong, Siriporn Phakdeepsaook</td>
<td>Kiyoko Takahashi</td>
</tr>
<tr>
<td></td>
<td>‘I’m a straight person. I’m not good at creating positive image’: reinvented self-images of Chuwit Kamolwisit in his 2004 campaign for the Bangkok governor</td>
<td>Allative preposition in Thai</td>
</tr>
<tr>
<td>16.30-17.00</td>
<td>Chanokporn Puapattanakun</td>
<td>Na Pham</td>
</tr>
<tr>
<td></td>
<td>Another Look at Thai Argumentative Discourse</td>
<td>Translating topic-comment structures of Vietnamese into English</td>
</tr>
<tr>
<td>17.15....</td>
<td>Book launch: &quot;The Tai Languages of Assam - a grammar and texts&quot; by Stephen Morey. Refreshments provided.</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Session 1 : Coombs Lecture Theatre</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| 9.30- 10.00 | Jamin Pelkey  
Dependent Clause Marking in Phowa |
| 10.00- 10.30 | Hsiu-chuan Liao  
Pronominal Forms in Central Cagayan Agta: Clitics or Agreement Features? |
| Break : refreshments in Coombs Lecture Theatre Foyer |
| 11.00- 11.30 | Nor Hashimah Jalaluddin  
Memper-Kan Verbs With Adjectival Roots In Malay: An Analysis From DBP Corpus |
| 11.30- 12.00 | Phil Quick  
A Lexicographical Introduction and Inventory of Pendau Fish Names |
| Lunch : Caterina's Cafe |
| 13.30- 17.00 | Excursion |
Editor’s preface

The 15th annual meeting of the Southeast Asian Linguistics Society (SEALSXV) was held at the Australian National University, from April 20-22, 2005. The meeting was hosted by the Department of Linguistics of the Research School of Pacific and Asian Studies and the Southeast Asia Centre of the Faculty of Asian Studies.

It was the first time that SEALS had met in Australia. Since its foundation by Martha Ratliff & Eric Schiller back in 1990 at Wayne State University (USA), SEALS has been held at various American and Asian institutions, not yet making it to other continents.

SEALS is an informal association that lacks a formal standing committee or fixed institutional home, although the Arizona State University Program for Southeast Asian Studies has published the proceedings and advertised them on a SEALS website at http://www.asu.edu/clas/pseas/seals.html. The “unorganised” organisation has worked quite well, with motivated individuals continuing to volunteer their time to organise meetings and to persuade their institutions to host them. However, at the time of writing, it appears that there will be something of a hiatus in the ongoing SEALS movement, and 2006 may pass by without a meeting. The good news is that Arizona State (the unofficial home of SEALS) may host a meeting in 2007.

Up to forty people attended the meeting in Canberra, and a wide range of papers were presented. As you can see from the contents, a number of the papers presented to the conference were not subsequently offered for the proceedings volume, but you may note their titles and presenters from the program. Also there are a couple of papers appearing here that were not actually read at the conference, as their presenters were unable to attend at the last moment, but I am delighted to be able to include them in this publication.

Unlike previous volumes of SEALS proceedings, this one is distributed electronically and free of charge, as a Pacific Linguistics e-book. Personally I am a strong advocate of electronic publication for academic work. These days the sheer cost of paper publication is becoming prohibitive, and libraries are cutting back on their subscriptions to hard copy periodicals and conference volumes. By contrast, e-books and e-journals are no more trouble to prepare than traditional publications, require no printing or inventory, and can be effortlessly distributed around the world, accessed by anyone with an internet connection at their home, work or local café. We now have the technical means to eliminate many publication costs and delays, allowing academia to catch up with what is now established best practice in many other industries.

Finally, I would like to express my gratitude for having the privilege to organise the SEALSXV meeting in Canberra, and thank the Department of Linguistics of the Research School of Pacific and Asian Studies and the Southeast Asia Centre of the Faculty of Asian Studies for their support. Such conferences are so important for academic life, and I feel honoured that so many people placed their trust in me to run the meeting. I thank all of the participants for their attendance and contributions, and hope that we will meet again at future gatherings.

Paul Sidwell
Canberra, October 2005.
0. Abstract
This paper concerns one type of relative clause in Indonesian, the variety of Malay which is the national language of Indonesia. This clause type has a verb marked by the meN- nasal prefix, and uses a gap strategy to relativise on an Object position – illustrated in (8) and (9) below. (In this paper the term ‘Object’ is used to refer collectively to both direct objects and objects of complements, as both are equally the focus of the paper. When it is necessary to distinguish the two, the terms ‘direct object’ or ‘object of a complement’ are used.)

This type of relative clause is rejected by grammarians and is apparently regarded as non-existent. This paper will demonstrate that it has a small but definite role in the formal variety of Indonesian. The paper will also identify factors that motivate its use, and explain why it is likely to become more common.

1. Background
1.1. Basic clause types
An active clause in Indonesian is as follows:

(1) Anak itu akan membantu Ali.
child that FUT MEN-help Ali.

That child will help Ali.

The word order is SVO. The verb is marked as active by the nasal prefix meN-. Two passive constructions exist in Indonesian, often called Passive Type 1 and Passive Type 2 (cf. Sneddon 1996). The choice between Type 1 and Type 2 is determined by who the Actor is.

Passive Type 1 is used when the Actor is third person.

(2) Ali akan dibantu (oleh) anak itu.
Ali FUT DI-help by child that

Ali will be helped by that child.

In (2), the verb bantu is marked as passive by the verb prefix di-. The Agent anak itu is placed after the verb, as an oblique constituent.

Passive Type 2 is used when the Agent is a pronoun or a pronoun substitute.

(3) Ali akan saya bantu.
Ali FUT 1sg ø-help

Ali will be helped by me.
1.2. Preverbal elements in passive clauses

A constraint on forming passive clauses is that certain lexical elements cannot always precede a passive verb. By way of illustration, let us first look at a word that can freely do so: the future marker akan. In (1) it precedes the active verb membantu ‘help’, while in (2) and (3) it precedes the passive verbs dibantu and bantu respectively, to form a passive clause that is semantically equivalent to the active clause.

A word that cannot precede a passive verb in that way, according to Wolff (1986) and some informants, is the word berani ‘dare.’ Here it is first in an active clause:

(4)   Ali berani melawan polisi itu.
Ali dare MEN-resist police that
Ali dared to resist those police.

In (4), it is the Agent, Ali, who is doing the ‘daring’. In other words, berani ‘dare’ is semantically linked to the Agent. But if we try to use berani that way in a passive clause, it is not accepted by all speakers:

(5)   ? Polisi itu berani dilawan oleh Ali
Police that dare DI-resist by Ali
For: Ali dared to resist those police.

Sentence (5) is intended to mean the same as (4), i.e. berani ‘dare’ is once again linked semantically to Ali. But when berani ‘dare’ precedes a passive verb while thus semantically linked to the Agent, some speakers regard the resulting clause as unacceptable.

It is hard to specify which words can and cannot precede a passive verb when semantically linked to the Agent. Both informants and writers disagree among each other in the case of certain words, with Alieva et al (1991: 390) contradicting Wolff (1986) by accepting berani ‘dare,’ and Alwi et al (2000:347-348) contradicting nearly everyone by ruling out ingin ‘wish, desire’ in some sentences. But the crucial thing for present purposes is simply that many words are problematic in this position. Ones that struck one or more of my informants as awkward to some degree include: berani ‘dare’, ikut ‘join in’, mampu ‘(be) capable’, berhasil ‘succeed’, gagal ‘fail’, suka ‘like’, lupa ‘forget, coba ‘try’, mulai ‘start’, takut ‘(be) afraid’, enggan ‘(be) reluctant’, senang ‘(be) happy’, malu ‘(be) ashamed’ and berusaha ‘try’.

1.3. Relative clauses

A subject relative clause (RC) in Indonesian is illustrated below. (In all RC examples both the relativiser yang and the position of the gapped noun phrase are highlighted in bold.)

(6)   Anak yang ø membantu Ali itu baik hati sekali.
child REL MEN-help Ali that kind very
The child who helped Ali is very kind.
Taboo object relative clauses in Indonesian

In (6), the relativised noun phrase (NP) is anak ‘child’. The position of the relativised NP is marked with a gap, indicated here by the null symbol ø. This NP is the subject of the verb membantu ‘correct,’ which is marked as active by the nasal prefix meN-. The relative clause is preceded by the relativising particle yang. This type of RC is often referred to as a gap strategy and is the most common type of RC in Indonesian (cf. Sneddon 1996, Cole & Hermon 1998).

According to most grammarians of Indonesian, one cannot relativise with the gap strategy on any position other than Subject. And when the RC verb is marked with a meN- prefix, then virtually all grammarians agree in ruling out any non-Subject position (e.g. Cole & Hermon (1998); Vamarasi (1999: 12-13); Kaswanti Purwo (1989: 350, 1996: 195); Musgrave (2001: 59); Sneddon (1996: 286); Voskuil (1996: 188-189)). Native-speaking informants also tend to regard such RCs as awkward to some degree, discussed later.

In line with this constraint, Indonesians routinely use passive RCs to ensure that a gapped NP occupies the Subject position. So a notion such as “the child whom Ali helped” will typically be rendered as:

(7) Anak yang ø dibantu oleh Ali itu kecil sekali.

The child whom Ali helped is very small.

The type of RC which is the focus of this paper disregards that constraint, and uses an active RC to gap an Object position instead. Here is an example, from a public address by the fiction writer Pramoedya Ananta Toer:

(8) Bisa saja dibuat daftar kekurangan atau kekeliruan angkatan-angkatan yang kalian perlu mengoreksi ø.

One could certainly make a list of the deficiencies or errors of the earliest generations (of freedom fighters) which you need to correct.

In (8), the relativised NP is kekurangan atau kekeliruan angkatan-angkatan terdahulu ‘deficiencies or errors of the earliest generations’. This NP is the direct object of the meN- prefixed verb mengoreksi ‘correct,’ and its position in the RC is marked with a gap.

A second example is this, from a writer’s personal weblog:
Isn’t our desire to create humans devoid of free will, a pointless desire…? So the mystery that we are not capable of grasping rationally is: why was humanity created?

In (9), the relativised NP is misteri ‘mystery’. This NP is the object of the men-prefixed verb menangkap ‘catch,’ and its position in the RC is marked with a gap. The lexical element mampu (‘be) capable’ is probably best regarded here as a full verb. As such, it is a matrix verb that takes the verb menangkap ‘catch’ as its verbal complement (see Wolff 1986: 142-144; Sneddon 1996: 270-271). By that analysis, the relativised NP misteri is not a direct object but rather the object of a complement.6

The types of relative clause in (8) and (9) – which I will refer to collectively as “men- gapped Object,” seem to be regarded as non-existent.7 The direct object variant receives more explicit attention, with Voskuil (1996: 203) stating flatly that instances of it “do not occur, in no variant of Indonesian”. Similarly, Ewing and Cumming (1998:79) say that the prescriptive ban on that clause type is closely adhered to in both formal discourse and more informal written texts. And while grammarians often fabricate instances of “men- gapped Object” for discussion, I have never seen an authentic instance of it cited by any writer.

The aims of the present study are: to demonstrate that the relative clause type “men- gapped Object” does occur in Indonesian; to describe the salient formal features of authentic tokens of this clause type and to identify factors motivating its use.

2. Method

The first stage of the research was casual observation. I simply recorded in a notebook, over several years, any instance of “men- gapped Object” that I happened to notice in any text (19 instances in total). The second stage was a small-scale search for additional instances of this construction in written texts on the internet, using Google search engine.8 This search consisted of six sessions of about two hours each. It was not intended to be exhaustive but rather to test out tentative notions formed on the basis of the initial observational data about what type of formal features tended to characterise tokens of “men- gapped Object,” and to give some idea of how frequently writers might use this clause type and in what type of texts.

3. Findings and Discussion

The study found that “men- gapped Object,” while very infrequent, is used in Indonesian and within the formal variety of Indonesian specifically. Nearly 100 Indonesian texts were found that contained an instance (or more than one) of this clause type.9 Of these texts, just under two-thirds were original Indonesian texts. The
Taboo object relative clauses in Indonesian

remainder were in texts that were translations or suspected translations of an English
text or, occasionally, of a text in Arabic or Dutch.\(^{10}\)

3.1. Register
Most instances of “meN- gapped Object” in the data occurred in texts that were formal
in register, i.e. marked by the consistent presence of overtly formal features and absence
of informal ones (see Ewing, to appear, for a description of informal Indonesian and
Sneddon (2003) for the relationship between the formal and informal varieties). As the
meN- nasal verb prefix itself is a marker of formal register to some degree, this is not
surprising. However some instances occurred in texts that were only semi-formal, and a
few occurred in quite strongly informal texts.\(^{11}\)

3.2. Types of texts containing “meN- gapped Object”
Tokens were found to occur in a range of text types. The tokens obtained by the initial
stage of casual observation included ones from television interviews, novels, short
stories, scholarly non-fiction works, and newspaper articles. The ones obtained by web
text-searching included ones written by journalists in articles in on-line newspapers and
magazines, and by a variety of people quoted or interviewed in such articles; as well as
entries in personal weblogs, postings to discussion forums, academic articles, essays on
social and religious issues, religious scriptures (both Christian and Muslim), short
stories, and transcripts of speeches by public figures such as government ministers.

These texts containing “meN- gapped Object” are nearly all very recent. Of
those whose date of composition could be ascertained, the vast majority were composed
in the 1990s or later. A couple were written in the 1970s (and first published in the
1980s), but none earlier than that.\(^{12}\)

3.3. Malaysian
The web text-searches revealed that this construction also occurs in Malaysian, the
variety of Malay that is the national language of Malaysia. This is significant as the
prescriptive taboo on this construction and the rejection of it by native informants is as
strong in Malaysian as in Indonesian (see Yeoh 1977).

Thirty Malaysian texts were found with at least one instance of “meN- gapped
Object.”\(^{13}\) Ten of these were original texts in Malaysian while the others were
translations or suspected translations of English (or Arabic) texts.

3.4. Influence of English: Translation from an English source text
Linguistic influence from English is a powerful motivation for the use of “meN- gapped
Object”. This is most obvious when writers produce it in the act of translating from a
source text in English. Here is an example:

\[(10)\quad \text{Peraturan semacam ini juga memaksakan seorang regulation one-type this also MEN-force-KAN one-CLASS}
\text{bawahan untuk memperbincangkan persoalan inferior COMP MEM-PER-discuss-KAN matter}
\text{pribadinya dengan para atasan dalam birokrasinya, personal-3sg with PL superior in bureaucracy-the}
\text{suatu situasi } \text{yang orang banyak pasti lebih suka} \]
This kind of rule (i.e. that civil servants have to ask their boss’s permission before they divorce or re-marry) also forces employees to discuss their personal affairs with their superiors in the bureaucracy, a situation which many people would definitely prefer to avoid.

In (10), the entire segment of head NP + relative clause (i.e. the segment suatu situasi yang orang banyak pasti lebih suka menghindari) is evidently a close translation of an English segment of text with an Object relative clause, such as “a situation which many people would definitely prefer to avoid”.\textsuperscript{14}

The Malaysian writer Muhammad Salleh is striking for his heavy use of “meN-gapped Object” (along with many features of gross translationese) in his Malaysian renderings from English of various texts on socialism. He wrote eight of the translated Malaysian texts in the data and he uses it liberally in all of them (e.g. in Salleh 2002).

3.5 Influence of English: Covert syntactic transfer
Speakers and writers who use “meN-gapped Object” in original texts, too, often do so at least partly because of syntactic transfer from English. The very fact that people use this construction when translating texts from English strongly suggests a covert influence at other times. Other factors also point to a powerful covert influence of English, as follows.

(a) English and Dutch have strongly influenced the syntax of Indonesian (see e.g. Moeliono 1992; Badudu 1996; Kaswanti Purwo 1996; Sneddon 2003). During recent decades the influence has increasingly been from English and, as Sneddon (1996: 2) observes, new structures are constantly being introduced under pressure from English. One such influence during recent decades has been a rise in frequency of active clauses as opposed to passive ones (Kaswanti Purwo 1996: 199-208; Verhaar 1989: 258-261). The emergence of this active construction of “meN-gapped Object” might be one manifestation of that broader syntactic influence.\textsuperscript{15}

(b) the users of “meN-gapped Object” in the data are well-educated Indonesians (e.g. journalists, fiction writers, university students or politicians). This social stratum is highly familiar with English, which is the first foreign language in Indonesia, a compulsory subject throughout high school and which carries very high prestige (see e.g. Sneddon 2003, Lowenberg 1994). In Malaysia English has greater prominence still, and educated Malaysians are generally more proficient in English than educated Indonesians.

(c) My informants often remarked without prompting that English influence was at work in original instances of “meN-gapped Object” in the data. Moreover, English native speaking students of Indonesian – in my own experience and that of other teachers – very often produce “meN-gapped Object” in their own speech and writing.
3.6 Formal features of “meN- gapped Object”

The study also revealed certain formal features of “meN- gapped Object” tokens in the data.

3.6.1 Preverbal lexical elements. Rather than comprising a minimal structure of obligatory elements only, most tokens of “meN- gapped Object” in the data contain lexical elements before the meN- prefixed verb such as negators, temporal/aspectuals, or modals. Fully 85% of the tokens found in original Indonesian texts contain extra words of this kind and a large majority of tokens in translations of texts also contain them.

An example is (8) earlier, where in front of the RC verb mengoreksi ‘to correct’, we find the modal auxiliary perlu ‘need’. Here are a few more examples to illustrate this tendency, with the additional lexical items marked in capitals: tidak ‘not’ in (11), tidak bisa ‘cannot’ in (12) and ingin ‘wish, desire’ in (13).

From a novel by the famous writer Pramoedya Ananta Toer:

(11) Sahaya hanya mengetahui yang orang Jawa TIDAK

1sg only MEN-know-i REL person Java NEG

mengetahui θ, karena pengetahuan itu milik bangsa

MEN-know-i because knowledge that possession nation

Eropa dan karena memang saya belajar dari mereka.

Europe and because indeed 1sg BER-study from them

It’s just that I know what the Javanese do not know, because that knowledge belongs to Europeans, and I have surely learned from them.

A well-known writer of fiction, Danarto, being interviewed in a newspaper article:

(12) Justru di dalam puisi Goenawan Mohamad itu ada

precisely LOC inside poetry Goenawan Mohamad that there-are

adegan senggama yang orang TIDAK BISA melihat θ.

scene intercourse REL person NEG can MEN-see.

Nah, dari sinilah …

Well from this-EMPH

In Goenawan Mohamad’s poetry for that matter, we find sex scenes which are not explicitly depicted. Now this is what…

And one example from Malaysian (a headless relative): in the transcript of a speech by then-President Mahathir Mohamad:

(13) Orang asing dan media mereka tidak ada niat baik

person foreign and media 3pl NEG have intention good

terhadap kita … Yang mereka INGIN melihat θ ialah negara

towards 1pl REL 3pl wish MEN-see COP country

kita ini mengalami nasib buruk …

1pl this MEN-experience-i fate bad

Foreigners and their media do not have friendly intentions towards us …

What they would like to see is our country suffering misfortune …
It would appear then that the presence of such words makes “meN- gapped Object” feel more acceptable to those who use it. Informants agreed that these extra words made the construction much more acceptable (although still tending to find it awkward to some degree).

This resembles the situation with a different type of Object relative clause in Indonesian: one that marks the object not with a gap but with a pronoun copy of the NP. With that construction, the addition of lexical elements such as modals improves it to the point of making it fully acceptable (Kaswanti Purwo 1984:145-147, 1989: 420; Verhaar 1983:60) – as illustrated below.

(14) Ini merupakan soal yang SUDAH LAMA kita INGIN membicarakan-nya.
    This MEN-form-KAN matter REL already 1pl wish MEN-talk-KAN-3sg
    This is a matter that we have wanted to discuss for a long time (from Verhaar 1983: 59)

Compare that with (15) below, stripped of extra words and thus unacceptable:

(15) *Ini merupakan soal yang kita membicarakan-nya.
    This MEN-form-KAN matter REL 1pl MEN-talk-KAN-3sg
    This is a matter that we discussed.

Why should extra words improve the type of RC in (15)? Kaswanti Purwo (1984) argues that they (a) create distance between the word yang and the meN- prefixed verb, and (b) make the clause in some sense ‘modal’, i.e. not purely factual and not referring to a definite event that has taken place. In this study, the same thing is probably happening with “meN- gapped Object”. So for example the added words in (11), (12) and (13), as well as obviously creating distance, also make the clause non-factual/ non-eventive.

3.6.2 Preverbal words that are awkward in passives. Many tokens of “meN- gapped Object” contain a lexical element that would be awkward in a passive clause. One type is a preverbal word of the kind discussed in section 1.2. Here are examples, with that preverbal element marked by capitals in each case: lebih suka ‘prefer’ in (16), gagal ‘fail’ in (17), ikut ‘join in’ in (18) and berani ‘dare’ in (19).

A radio interviewer’s question to a contestant in the Science Olympics:

(16) Kalau Ina bagaimana? Apa ada bagian dari ilmu-ilmu sains yang Ina LEBIH SUKA menekuni 8?
    If Ina how INT there-is part from knowledge-knowledge science REL Ina more like MEN-persevere-i
    And what about you, Ina? Is there an area of science that you particularly like to study in depth?
Taboo object relative clauses in Indonesian

From a short story by the writer Pramoedya Ananta Toer:

(17) “...sangat indah, usia muda, semua bisa menangani,” extremely beautiful age young all can MEN-handle-i ia mengangguk-angguk ria, “bagus, teruskan apa yang he MEN-nod-nod happy fine continue-KAN what REL telah kami GAGAL melakukan ø.17 Dulu. already 1pl fail MEN-do-KAN formerly it’s such a beautiful thing to be young, you can do anything”, he said nodding happily. “That’s good: carry on with the task that we failed to accomplish. In the old days.”

From a newspaper report about the arrest of several thieves (a headless relative clause):

(18) ... ketiga tersangka terakhir ini adalah residivis ... Selain the-three suspect last this COP recidivist apart-from ketiga kasus tadi, kasus apalagi yang anda IKUT the-three case just-now case what-else REL 2sg join-in merancang ø? tanya Lettu Irwan Anwar. MEN-plan ask first-lieutenant Irwan Anwar ... these three suspects are repeat offenders. ... “Apart from the three offences we’ve just mentioned, what other ones have you helped to plan?” asked First Lieutenant Irwan Anwar.

From a pop novel by a well-known author (a headless relative clause). [A young man, Fadila, likes to tease his younger sister’s female friends.]

(19) Cuma Saskia yang Fadila tidak BERANI mengusik ø. only Saskia REL Fad ila NEG dare MEN-tease Saskia was the only one who Fadila didn’t dare to tease.

In each of 16-19 above, an equivalent passive RC would be awkward because the highlighted words could not link semantically to the Agent without creating a clumsy effect (see 1.2).18 Instances of “meN- gapped Object” containing such lexical items were common in the data. Of the tokens obtained initially by field observation, seven of the 19 had such an element (lebih suka ‘prefer’ (x 2), gagal ‘fail’, berani ‘dare’, lupa ‘forget’, coba ‘try’, and ikut ‘join in’). And the Google text-searches produced multiple instances with each of mampu ‘(be) capable’, ‘lebih suka ‘prefer’, coba ‘try’, ikut ‘join in’, gagal ‘fail’ and senang ‘(be) happy’, plus single tokens with malu ‘(be) ashamed’, mulai ‘start’, takut ‘(be) afraid’, enggan ‘(be) reluctant’ and berusaha ‘try’.

The presence of these preverbal words makes the clause sound more acceptable to informants (although seldom entirely so). These words have the same facilitating effect as the preverbal words in 3.6.1, i.e. they create distance between yang and the meN- prefixed verb and often make the clause non-factual and/or non-eventive as well. But these particular ones also create a constraint on using a passive clause – another motivation for using “meN- gapped Object” instead in their presence.19
3.6.3. Pronoun modifiers which are awkward in passives. Instances of “meN- gapped Object” often contain an in situ modifier of a pronoun Agent phrase. A prime example is the modifying word sendiri (‘self’). Here is such an example, from a short story by a respected writer:

(20) Tahun yang lalu rekan-rekan wartawan Roma
year REL past colleague-colleague journalist Roma
telah memilihnya sebagai anggota yang paling
already MEN-choose-3sg as member REL most
perlente busananya – suatu pujian yang dia SENDIRI
stylish fashion-3sg a praise REL 3sg self
tidak pernah memimpikan ø.
NEG ever MEN-dream-KAN

The previous year Roma’s journalist colleagues had voted him the most stylishly dressed among them – praise that he personally had never dreamed of.

In (20), the pronoun Agent is modified by sendiri ‘self’, to make an Agent phrase meaning ‘he personally; he himself’. Other in situ modifiers of Agent pronouns in the data were the quantifiers semua ‘all’ and sama-sama ‘both, all’ (e.g. in the headless relative yang kita SEMUA menyaksikan ø ‘what we ALL witness,’ and in organisasi yang kita SAMA-SAMA menganggotai ø ‘the organisation which we ALL belong to’).

This feature probably tends to occur in the active, “meN- gapped Object” construction because it is awkward to use in a passive clause. With Passive Type 2 specifically, the inclusion of a pronoun Agent phrase modified in situ by any other element is doubtful or unacceptable (cf. Musgrave 2001: 84). So for example, a passive RC such as (21) would be ruled out by many speakers:

(21) ? suatu pujian yang tidak pernah dia SENDIRI mimpikan ø.
a praise REL NEG ever 3sg self ø-dream-KAN
praise that he personally had never dreamed of (literally: ‘praise that had never been dreamed of by him himself’)

To make (21) fully acceptable, the element sendiri ‘self’ intervening between the Agent pronoun dia and the verb mimpikan must be removed.

Altogether, 23 tokens of “meN- gapped Object” contained an in-situ modifier of the Agent phrase. Nearly 20% of the tokens in original Indonesian texts contained one, usually sendiri ‘self’. This should not be taken as a valid measure of proportions as some of my web text-searching specifically targeted these modifiers, but does indicate that this is a likely formal environment for “meN- gapped Object” to occur in.

These Agent modifiers made the data examples more acceptable to informants, and made some examples wholly acceptable to one informant. As well as simply creating distance between yang and the meN- prefixed verb (see above), they probably make this active construction sound better by ruling out a ready passive alternative.
3.7 Non-restrictive relatives. A considerable number of “meN- gapped Object” instances were non-restrictive relative clauses, i.e. ones which merely add extra information about an entity rather than identifying that entity within a larger set.22 Fifteen of the Indonesian texts with “meN- gapped Object” contained a non-restrictive token (or more than one), and several of the Malaysian texts contained one or more such tokens as well. A few non-restrictive examples of “meN- gapped Object” from original Indonesian texts:

In a novel by the writer Pramoedya Ananta Toer:

(22) Ya, Multatuli, di samping Domine Baron von Hoëvell itu, dan seorang lagi, yang barangkali saja gurumu lupa menyampaikan ø, yakni Roorda van Eysinga.

Yes, Multatuli, besides Domine Baron von Hoëvell, and one other, who perhaps your teacher forgot to mention, namely Roorda van Eysinga.

A politician quoted in a magazine article:

(23) Pokok-pokok Deklarasi Rio sudah dicantumkan. Tidak mungkin kita melanggar perjanjian internasional, yang kita ikut meratifikasi ø”, kata Walanrugian.

The basic points of the Rio Declaration have already been laid out. It isn’t possible that we would violate an international treaty, which we also have ratified”, said Walanrugian.

A double-instance in a posting to a discussion forum on the topic of autism:

(24) Yang terakhir yang saya sendiri belum bisa melakukan ø dengan penuh dan sepanjang waktu, dan akan tetap berusaha untuk bisa melakukan ø, ikhlasakan kondisi tersebut, karena semata-mata kehendak Alloh SWT...

The last thing I would suggest, which I myself am unable to do fully and consistently, and will keep trying to be able to do, is to accept that situation willingly, because it is entirely the will of Almighty God.

The non-restrictive nature of such tokens seems to make them more acceptable. For instance, (22) was relatively acceptable to informants, and fully acceptable to one, who remarked that “it’s because of the commas – they make it sound fine”. But we
should note that examples such as (22) – (24) above contain other features as well which make them sound smoother: extra lexical elements before the *meN*-prefixed verb in each instance, plus the adverbial *barangkali saja* ‘possibly’ in (22) and the Agent modifier *sendiri* ‘self’ in the first of the tokens of (24). So it is uncertain how much of a part the non-restrictive nature of the clause actually plays.

3.8 Individual variation. The last thing to mention is the issue of idiolect and personal variation. Seven of the tokens in the data were by the one writer: Pramoedya Ananta Toer, Indonesia’s most internationally famous writer of fiction. The instances (8), (11), (17) and (22) are all from works of his, and his seven tokens in the data were from five separate works: two novels, an autobiography, a short story and a public address. So while we can assume that many educated Indonesians never use this construction, the data suggest that others may use it with some regularity as a feature of their idiolect.

4. Conclusion
This study shows that that gapped object relative clauses do occur in Indonesian in the presence of the *meN*-verb prefix. While their textual frequency is extremely low and they are still generally regarded as ungrammatical, they have a marginal role in the formal variety of Indonesian.

A major reason people use this clause type is influence of English. Sometimes the influence of English is overt, in the form of translation from an English source text. Often it is more covert, in the form of second language syntactic transfer.

This clause type has some striking formal features. One is the frequent presence of preverbal lexical elements such as negators, temporal/aspectual markers or modals. Another is the common inclusion of elements that make it harder to use a passive clause instead. These include preverbal lexical elements like *mampu* ‘(be) capable’ and *berani* ‘dare,’ as well as *in situ* modifiers of pronoun Agents, such as *sendiri* ‘self’. A final observation is that this type of RC is sometimes non-restrictive.

The factors above probably interact in many cases. For instance transfer from English probably interacts with the presence of lexical elements conducive to this type of RC, to motivate speakers/writers to use it.

Individual preference also seems important, as one famous writer employs this construction as a small but definite part of his idiolect.

Finally, the study extends our knowledge of English influence on Indonesian by identifying this structure as one for which English is largely responsible. As English influence on the language keeps growing (at least for the foreseeable future), this type of clause will probably become more widely and heavily used.

Notes

1 I am grateful to the following Indonesians who helped with this study by supplying judgements on the acceptability of many data examples: Amrih Widodo, Catharina Williams, Djasamen Saragih, Hamdan Juhaniss and Urip Sutyono; as well as to Ben Arps for discussing Dutch relative clauses with me.

2 *meN*- is the prefix of a great many active transitive verbs. Voskuil (1996: 61) regards *meN*- as a marker of active voice. Musgrave (2001: 102) says that while it does not mark active voice “in the sense usually understood”, it does signal that the Actor of the clause is linked to the Subject function. Cole and Hermon (1998:2) call it a transitive prefix. The fact that *meN*- also
prefixes a small number of intransitive verbs makes it hard to characterise its function in a simple way.

3 Wolff (1986:143-144) and Vamarasi (1999: 141-146) claim that it is full verbs which cannot precede a passive verb when linked semantically to the Agent. Sneddon (1996: 270-271) offers a slightly different analysis, implying that a few words can precede a passive verb in that way and yet still function as verbs while so doing (namely berhasil ‘succeed’, gagal ‘fail’, mau ‘want’ and ingin ‘wish, desire’).

4 Keenan (1972) and Keenan and Comrie (1977) both claim that gapped relativisation of the direct object is possible in the presence of the “meN-” prefix, but the fabricated example they offer is explicitly rejected by both Musgrave (2001: 240) and Yeoh (1977: 86), and Comrie himself later says that it is not possible to relativise on the direct object in Indonesian (Comrie 1989: 157). Chung (1976) implicitly accepts “meN- gapped Object” with objects of complements – see note 6 below. In the absence of a meN- prefix on the verb, certain grammarians (Voskuil 1996; Musgrave 2001; Cole & Hermon 1998) accept gapped relatives on the direct object (or ‘non-subject argument’ as two of them prefer to call it). They all acknowledge that this RC type, as well, is prescriptively frowned upon.

5 I have not marked instances of this clause type with a * or ? symbol in this paper, as all cited instances are authentic textual examples.

6 Whether a preverbal lexical element in an active clause is a full verb can be hard to determine. A primary test is whether that lexical element can precede the verb in an equivalent passive clause: auxiliary words can do so while full verbs cannot (Wolff 1986: 142-143; Vamarasi 1999: 142). Another test is whether the complementiser untuk can intervene between that lexical element and the main verb in the active clause (Vamarasi 1999: 143). However both these tests yield an equivocal answer for many lexical elements. Chung (1976: 51) seems to accept “meN- gapped Object” when the relativised NP is the object of a complement rather than a direct object. She does not say so outright, but proffers two (fabricated) examples of gapped non-subject relative clauses that she finds acceptable and contends that they prove how meN- prefix can appear on the verb under “ill-understood conditions”. The gapped NP in one instance appears to be the object of a sentence complement, and in the other instance, the object of a complement verb (see examples 36a and 36b of Chung’s text). Unfortunately no other writers pass a direct judgement on Chung’s two sentences. Cole, Hermon & Tjung (2003) claim that Musgrave (2001) explicitly rejects both, but my own reading of Musgrave (2001) found no mention of either.

7 Cole, Hermon and Tjung (2003) find that in colloquial Jakartan Indonesian, gapped object relatives occasionally occur in speech in the presence of the nasal verb prefix N- (not meN-). They argue that this happens because the nasal N- prefix in Jakartan Indonesian is losing its status as a marker of active transitive verbs (and as support for that latter assertion, cite a 2002 paper by David Gil).

8 I chose an internet search-engine rather than concordance software such as WordSmith Tools because it would have taken too long to create a large enough corpus to search using the latter. The Google search engine allowed (somewhat clumsy) searching of an enormous corpus of text. Its main disadvantage was that it does not allow truncation of words, so I could not search a given slot for the presence of a meN- prefixed verb by entering only the letters “me”. The few web search engines that do allow truncation (e.g. Alta Vista) proved to have other disadvantages that outweighed it. A typical Google search entry started with the relativiser yang and looked something like this: “yang saya OR aku OR kamu OR Anda OR dia OR ia OR mereka mampu OR gagal OR suka OR berhasil OR berani.” In some searches the ‘wild card function’ of the asterisk was also useful, i.e. an asterisk entered as a search term yields any word occurring in that slot. As the above example of a search entry shows, my search terms tended to specify pronoun Agents rather than full noun Agents because of the conveniently small number of the former. This naturally biased the data results towards tokens containing that type of Agent.
Shortly after presenting this paper I chanced upon a much faster way to locate instances of “meN- gapped Object” in web-based texts, namely by specifying common meN- prefixed verbs as terms within the Google search entries, e.g. membeli ‘to buy’, mengalami ‘to experience’, menjawab ‘to answer’ or melihat ‘to see’. By brief experimentation with that new search technique I found another forty texts containing ‘meN- gapped Object’ – not analysed in this paper.

In the case of unattributed and possibly old translations of extracts from the Bible, the source text was quite likely Dutch. In the case of quotations from Islamic scriptures, it was quite likely often Arabic. Dutch Object relative clauses are closely parallel in structure to “meN-gapped Object” in much the same way as English ones are, and as far as I can see, Arabic Object relative clauses are also closely parallel to it when their optional resumptive pronoun is absent (cf. Gensler 2004).

I probably skewed my web-search results towards tokens occurring in formal register by my choice of search terms, to some extent. For example, when specifying lexical items for the pronoun slot I tended to leave out the distinctly colloquial 1st person pronoun form gue/gua.

In the case of quotations from translated editions of the Bible or the Koran, generally no information was provided about what edition was being quoted from or when the translation was written.

Despite the smaller number of web-text instances in Malaysian, these reveal a degree of penetration by “meN-gapped Object” into not only the formal but the official language. I found three tokens in separate speeches by then-President Mahathir Mohamad and one in a speech by a High Court Judge on his auguration, two instances on a government web-site giving information for applicants for a civil service examination, and two in a statute concerning the surrender of library materials to the government.

The structural parallels between English Object relative clauses and “meN-gapped Object” are very close. Just as the head NP in the English construction is followed by a relative pronoun who or which or complementiser that, the head NP in “meN-gapped Object” is followed by the relativiser yang. And the order of subsequent elements within the RC is the same in both cases: Agent phrase + optional modifiers/ matrix verb + active verb. (In the case of English the element who/ which/that is actually optional, e.g. “the book (that) I bought yesterday”; I don’t know how that fact affects perceptions of structural similarity.)

Another active RC type has also appeared in recent decades under influence of English. This is a gapped object RC without meN- prefix on the verb, such as yang saya tak harapkan ø ‘which I do not expect’ (Kaswanti Purwo 1996: 195; plus see p 208 for attribution to English influence). While Kaswanti Purwo (1996) himself implicitly regards that RC type as a passive clause, Voskuil (1996), Musgrave (2001) and Cole & Hermon (1998) all convincingly analyse it as active. Other forms of English influence on Indonesian syntax include the emergence of RCs employing relative pronouns (Moeliono 1992: 31; Badudu 1996: 36), the appearance of formal equivalents to English participle constructions (Moeliono 1992: 31; Kaswanti Purwo 1996: 196-197), and the rise in the use of copulas (e.g. Kaswanti Purwo 1996:192-193; Badudu 1996: 35).

Kaswanti Purwo (1984: 147) actually refers to distance between yang and the subject of the relative clause (“konstituen subjek dalam klausa yang bersangkutan”), but I cannot fully see how that formulation would apply in all of his own data examples.

Interestingly, in the very next line of this story the interlocutor echoes this statement by using gagal in a passive clause instead: Apa yang ø telah gagal Bapak lakukan? ‘What is it that you failed to carry out?’ (literally, ‘that failed to be carried out by you?’). Some informants judged this reply to be more awkward than the “meN-gapped Object” construction in (17) that preceded it.

In the case of (19) specifically, the equivalent passive clause would also run a high risk of being misunderstood. Such a passive clause, i.e. Cuma Saskia yang ø tidak berani diusik (oleh) Fadila, might be taken to mean “Only Saskia didn’t dare to be teased by Fadila”. That
Taboo object relative clauses in Indonesian

is, readers would interpret the lexical element berani ‘dare’ in the passive clause as linking semantically to the Undergoer ‘Saskia’. So in (19) “men- gapped Object” functions partly to avoid semantic ambiguity.

19 The fact that these preverbal elements can be analysed as verbs themselves (see 1.2) is significant. It constitutes another parallel with the pronoun-copying type of Object RC – since another way to render that construction acceptable is to insert a matrix verb (Kaswanti Purwo 1989: 420 f3) or a clause (Kaswanti Purwo 1984: 147).

20 The word sama-sama is most often used as a comitative meaning ‘together’ and in that sense it modifies the verb or entire clause, but it can mean ‘both, all’ (see Stevens & Schmidgall-Tellings 2004), and in that latter sense it modifies the Agent specifically. Stevens & Schmidgall-Tellings (2004) gloss sama-sama in this latter use as colloquial in register, but the texts in which it occurred in these data were not noticeably informal.

21 There are often ways around this constraint. One can float certain modifiers of the Agent to a position after the verb instead. This can be done with the modifier sendiri ‘self’ – although not with the quantifier semua ‘all’ (see Musgrave 2001: 184) nor, it seems to me, with the quantifier sama-sama ‘both, ‘all’ without changing it into a comitative word meaning ‘together’. And if the pronoun Agent is third person specifically, one could use a Passive Type 1 instead of Type 2. But the data nevertheless suggest that a desire to modify the pronoun Agent in situ is a motivation for using the “men- gapped Object” relative clause.

22 Keenan (1985) confines his notion of a relative clause to restrictive RCs only, calling non-restrictive ones ‘relative-like’ structures (Keenan 1985: 168-169). Little is written about non-restrictive RCs in Indonesian. Sneddon (1996) does not discuss them except for one special class that has marginal status (Sneddon 1996: 288: section 3.101). While Alwi et al (2000: 391) include one non-restrictive RC among their list of RC examples they do not remark on its difference from the others. In written Indonesian non-restrictive RCs are often set apart from the rest of the sentence by commas, although this convention is not always observed.

List of abbreviations in glosses
1sg, 2sg, 3sg 1st, 2nd, 3rd person singular
1pl, 2pl, 3pl 1st, 2nd, 3rd person plural
CLASS classifier
COMP complementiser
COP copula
EMPH emphatic marker
FUT future temporal marker
INT interrogative particle
LOC locative marker
NEG negator
PL plural marker
PROG progressive aspect marker
REL relativising particle

Appendix: Source of numbered data examples from authentic texts


(9) Webpage titled ‘Pertanyaan-pertanyaan dasar manusia’. Personal weblog of an unnamed writer (from country-specific references in text evidently an Indonesian writing around the year 2000). http://brasouw.tripod.com/isi/manusia_diciptakan.htm (accessed 30/1/05)
References
Taboo object relative clauses in Indonesian


THE INVISIBLE AGENT WITH GLOBAL MEANING:
THAI ZERO ANAPHOR SUBJECTS

Charatdao Intrarat
Department of Language, School of Liberal Arts,
King Mongkut’s University of Technology
charatdao20003@yahoo.com

1. Introduction
In global communication, the insight of a language enriches the mutual understanding between foreign and native speakers. When compared with English, Southeast Asian languages possess several distinctive features that usually cause misleading, if not confusing, interpretation. The frequent occurrence of zero anaphora in Cambodian, Malay and especially Thai languages, is one dominantly typical characteristic.

Usually, when a speaker assumes that the listener shares with him the same understanding or background knowledge, the noun phrase (NP) that refers to this ‘given information’ is omitted and appears in the sentence as a zero pronoun which is ‘invisible’ but still meaningful.

Compared with the English counterpart, zero anaphora occurs more frequently in Thai and can be classified into four types according to syntactic functions: SUBJECT, DIRECT OBJECT, INDIRECT OBJECT, and MODIFIER and four types of semantic roles: AGENTIVE, PATIENT, DATIVE/ BENEFACTIVE and GENITIVE and it outnumber the English counterparts (Intrarat 2003). No wonder this invisible agent with global meaning naturally causes problems to English speakers studying Thai and Thai students studying English alike.

Development of zero anaphora study
We can find different terms such as ZERO ANAPHORA, ZERO ANAPHOR, ZERO PRONOUN, NULL ELEMENT and EMPTY CATEGORY. All of them refer to the same constituent but from different viewpoints. ZERO ANAPHOR or ZERO PRONOUN refers to the pronoun that is omitted from a sentence. ZERO ANAPHORA is the general term that focuses on the relation in which that pronoun is seen as linked to an antecedent;Null ELEMENT is mentioned in Generative Grammar when we talk about parts of speech such as null subject; EMPTY CATEGORY is named for a gap in the structure as analyzed in Government and Binding Theory.

Zero anaphora has attracted interest from quite a few linguists for more than two decades. They started with Generative Grammar. Some Thai linguists for example Surintramont (1979), Kob siriphat (1988) and Hooncharmong (1991) worked with the process of analyzing and identifying antecedents for zero anaphora, applying Government and Binding Theory, but especially Control Theory.

Other linguists turn away from Generative Grammar to Pragmatics Discourse framework. Fox (1993) provides an account of distribution of a subset of anaphora in English conversation and expository prose. Geluykens (1994) also shows evidence that the distribution of anaphora in English conversation can be determined by the interaction of two pragmatic principles: the Clarity Principle and the Economy Principle. Huang (1995) observes that a single or a few syntactic parameters may never be

---

Paul Sidwell, ed. SEAUXV: papers from the 15th meeting of the Southeast Asian Linguistics Society.
adequate in accounting for null subjects and null objects in several languages. He states that different languages may require different licensing and identification strategies, some of which are clearly pragmatic/discourse in nature. In his cross-linguistic study of 550 languages, Huang (2000) provides an account of binding theory and control theory as well as various other syntactic, semantic, and pragmatic approaches to anaphora. Gutman (2004) states further that only a detailed theory of discourse anaphora can account for the distribution of third-person pro-drop in Hebrew, Finnish and Rumanian languages.

In Functional Grammar framework, Givon (1995) studies about zero anaphora and compares the occurrence of this device in different languages. He finds that the majority of sentence subjects in Ute, English, and two English-based pidgins: Spanglish and Filipinglish, have more zero anaphor or pronouns rather than the full NPs. Shokouhi (1996) reports that the occurrences in full NP form are more numerous in Persian than in English due to the use of formulaic politeness expressions. Although zero anaphora has been found to be pervasive in Persian conversation, it is constrained by factors such as turn-taking, repetition, and other interactional features.

Interpretation and identification of zero anaphora

In languages other than English, Huang (1994) focuses on the pragmatic aspects of interpreting several types of anaphors, including zero anaphors in Chinese. He also presents a revised neo-Gricean pragmatic theory of anaphora (Huang 2000). That is, the interpretation of certain patterns of anaphora can be made using general pragmatic inference, depending on the language user’s knowledge of the range of options available in the grammar, and of the systematic use or avoidance of particular linguistic expressions or structures on particular occasions. Matsas (1997) focusing on spoken Kurmanji, a northwestern Iranian language, identifies tentative restrictions on zero anaphora in conjoined clauses.

In Thai, Aroonmanakun (1997) and (2000) studies zero anaphora in the approach of Natural Language Processing and assumes that the resolution of zero pronouns can be done at two levels: the sentence level and the discourse level. The resolution at the sentence level can be implemented on the basis of principles in a sentence grammar, which is in accordance with the Government and Binding Theory. Zero pronouns that cannot be resolved by the Government and Binding Theory are resolved on the basis of discourse principles.

In cognitive aspects, Long and De Ley (2000) are interested in the antecedent of an anaphor in dialogues and how the antecedent becomes the most activated entity in the reader’s discourse model. They suggest that readers use knowledge about the structure of dialogue to anticipate the antecedent of an upcoming anaphor.

From the syntactic-semantic viewpoint, this researcher investigated in 1994 the recoverability of the zero anaphora in conjoined and complex sentences. I found that the antecedent of zero anaphora in Thai can be recovered within three conditions as follows:

1. Syntactic condition: Zero anaphora may refer to specific head nouns in specific sentence structures.
2. Semantic and context condition: Zero anaphora may refer to specific head nouns according to the selectional restrictions between the verbs of zero anaphora and those probable head nouns.
Thai zero anaphor subjects

3. Implicit condition: Zero anaphora does not always refer to the closest noun phrase or to the noun phrases recoverable from the immediate context. The reader has to search the antecedent from the extra-linguistic context, from background knowledge or from pragmatic rules.

From my study (Intrarat 2003), which is based on functional and cognitive approaches, this researcher observed that Thai zero anaphora performs more syntactic functions and semantic roles than their English counterparts. It is clearly seen that Thai zero anaphora may exist without any explicit antecedent in immediate context. It may refer to the antecedent at a greater distance later. Moreover, in some lyrics used as data, the antecedents are omitted throughout the whole lyrics. Due to the requirement of explicit antecedents in immediate context, English zero anaphora can not occur continuously but refer to different antecedents although this occurrence is acceptable in Thai.

From the study mentioned above, zero anaphora in Thai occurs most frequently in context of song lyrics when compared with pronouns and full NPs. It constitutes 57.05 per cent of the total occurrence of the three elements in the data. It is also remarkable that Thai zero anaphors can perform several semantic roles as well as pronouns or full NPs. The majority of zero anaphors occurs as the subject of the sentence and performs an AGENTIVE role. This is in the same trend as in other languages studied by Givon (1995) and Li (1997), but the difference is that the antecedent of the zero anaphora in Thai may not be explicit in the immediate context. Sometimes, this is implicit throughout the lyrics.

Since the majority of Thai zero anaphors occurs as the subject of the sentence, it would be interesting to investigate further to what extent they could be interpreted and what type of antecedent they could be identified as. In this study, the types of antecedents of zero anaphor subjects are analyzed and compared with English counterparts.

2. Purpose of the study
This article reports the study on the types of antecedents of Thai zero anaphor subjects and compared with their English counterparts. This analysis is aimed at supplying more background knowledge to English speakers who are studying Thai and making Thai students aware of their mother tongue interference in dropping SUBJECT when writing English. It is also aimed at reminding English teachers in designing English supplementary and remedial material for Thai students.

3. Hypothesis
This study set a hypothesis that Thai zero anaphor subjects can refer to more antecedents than their English counterparts.

4. Data
The data used in this study came from 5 contemporary newspapers and magazines namely, Matichon, Art and Culture Magazine, Sen Thang Sethee, Gourmet and Cuisine, and Tourism Authority of Thailand Magazine. Ten articles were randomly collected from each source from 1992 to 2003. The articles were studied and analyzed for zero anaphora occurring as the subject of a verb. These zero anaphors are further analyzed
for the identification of their antecedents. Then they are classified and compared with equivalent English counterparts.

5. Findings
From the data, I found that when zero anaphora occurred as SUBJECT and performed AGENTIVE role, there are 4 types of antecedents as follows:

1. First-person SUBJECT
2. Second-person SUBJECT
3. Third-person singular/plural SUBJECT
4. First/second/third-person/all-person SUBJECT

Finding 1 Zero anaphora as first-person SUBJECT.
For example

1.1) Thai: วันนี้ฉันเล่าเรื่องราวที่เกิดขึ้นมาครับ
   English: ‘Today I tell good story about young people once’

1.2) Thai: ฉันเชื่อว่าทุกคนไม่เคยมา
   องค์พิจารณา
   English: ‘I believe that many people have never been there’

In examples 1.1) and 1.2), zero anaphors occur as the SUBJECT of the verb ‘tell’ and ‘believe’ respectively. These subjects refer to the writer himself/herself so the subjects in these two sentences are first-person singular.

Finding 2 Zero anaphora as second-person SUBJECT
For example

2.1) Thai: ถ้าคุณไม่ทำให้คุณเข้ารูป
   องค์พิจารณา
   English: ‘If you argue that you can go by plane...’

2.2) Thai: ถ้าคุณอยากได้เที่ยว
   องค์พิจารณา
   English: ‘You can fly directly from Bangkok. You’ll have no difficulty’.

In these above examples, zero anaphors occur as SUBJECT of the verb ‘argue’, ‘go’ ‘fly’ and ‘see’. They refer to the particular person who is reading at that moment therefore these zeros are second-person SUBJECTS.
Finding 3  Zero anaphora as third-person singular/plural SUBJECT.
For example

3.1) Thai: สิ่งจำเป็น...ให้ทุกคนสังเกตเห็นผู้เป็นคนดีเรียกตัวอย่าง และจะมีนักโทษหรือผู้ต้องขัง
The boy...tell the see condition addicts in community then feel pity because are disgusted by society.

English: ‘The boy told me that he felt pity when he saw addicts’ condition in the community because they were disgusted by society’.

3.2) Thai: องค์การผู้ปกครอง...เมื่อผู้คนถูกขู่ขู่ให้เป็นไปตามที่ต้องการ
We may imagine people in long queue...when reach the turn able to step into to the well-lit shop laid with carpet so soft that can almost lie down, will have what happiness

English: ‘You may imagine people in long queue...when they reach their turn and step into the well-lit shop laid with carpet so soft that they car. almost lie down, how would they feel happy?’

In examples 3.1) and 3.2), it is noticeable that zero anaphors occur at several positions in one sentence and they refer to different persons. This particular characteristic is different from English zero anaphor which refers to one antecedent in one sentence.

In 3.1), the first zero anaphor occurs as the OBJECT of the verb ‘tell’ and performs the PATIENT role but this is not the focus of this present study. Those who are interested in details about the semantic roles of zero anaphora in Thai may read about them in my previous investigation (Intratat 2003).

The second and third zero anaphors in 3.1) refer to the boy, so they are third-person singular SUBJECTS of the verbs ‘see’ and ‘feel’. The last zero anaphor refers to the addicts who were disgusted, so this is the third-person plural SUBJECT of the verb ‘disgust’ and performs the PATIENT role.

In example 3.2), the first zero anaphor refers to the reader so it is the second-person SUBJECT of the verb ‘imagine’. The four zero anaphors that follow in the sentence refer to the people who have been waiting in a long queue. They are the third-person plural SUBJECTS of the verbs ‘reach’, ‘step’, ‘lie’ and ‘feel’.

Finding 4  Zero anaphora as first/ second/ third-person or all-person SUBJECTS.
For example

4.1) Thai: แต่งเนื่องกับอุตสาหกรรม การสร้างข้อหักขึ้นไม่เป็นไปตามที่ได้รับความรับ
nowadays any food eat not make fat not contains any fat will receive popularity

English: ‘Nowadays, any food you/ they/ whoever eats, that doesn’t make me/you/ them/ whomever fit and doesn’t contain fat, will be popular’.

23
4.2) **Thai:** ตัวชี้นำแกรนใส่ตัวหนังสือ...ถ้าลงที่ด้านซ้ายและใช้วิธีอื่นจะจารึกกลายเป็น ดังนั้นชี้ไปกับบริเวณ
When ø turn left, ø look right...ø must see and hear motorcycles roaring go all the area.
**English:** ‘When I/ you/ they/ whoever turns to the left or looks to the right,
I/ you/ they/ whoever must see and hear motorcycles roaring all over the area.’

In the above examples, zero anaphors occur as SUBJECTS of the verbs ‘eat’, ‘make’ and ‘contain’ in 4.1) and the verbs ‘turn’, ‘look’ and ‘see’ in 4.2). They can refer to the writer himself, the reader or someone in general, whoever it is, any food or consumers in 4.1 or pedestrians in 4.2). These zero anaphors refer to the first/second/third persons or all-persons. The antecedent here is sometimes ambiguous and can be interpreted differently and subjectively. Nevertheless, it is acceptable in Thai though not in English.

Statistics of the findings
From the total occurrence of entire zero anaphora in the data, the group with third-person singular and plural antecedents occurs the most frequently, which constitutes 55.07%. The second in frequency is the all-person antecedent type at 29.88%, the third is the first-person antecedent type at 11.52%. The least frequent group is the second-person antecedent type, which constitutes only 3.6%. The pie chart of this statistics is shown below:

**Chart 1: Antecedents of Thai Zero Anaphor Subjects**

Interpretation of the findings
From the statistics, most Thai zero anaphor subjects in the data refer to third-person singular and plural antecedents. It can be interpreted that this phenomenon is natural since the data came from narrative types of writing. Because it is the one-way communication from a writer to the readers, the ‘given information’ then mostly concerns the third party. Consequently, the zero anaphors referring to third-person antecedents frequently occurs. Different phenomenon may occur in face-to-face conversations, dialogues, letters and songs when the speaker corresponds directly with the listener or seems to do as such. In this situation, the first-person and the second
Thai zero anaphor subjects

person are not necessarily mentioned and tend to be omitted as zero anaphors referring to first and second-person antecedents more than other types.

As for the all-person antecedent, the second highest percentage type in this study, it can be interpreted that in some informal writing, the writer may want to construct a friendly atmosphere or to convince the readers to accept his ideas. He would appear to be less detached from the readers when he uses zero anaphora referring to all-person antecedents because he seems to integrate into the same group as the readers or whoever in general. It is also possible that, according to Thai social etiquette, the proper address terms to individual persons cause difficulties for the writer because he doesn’t know who his readers will be, so zero anaphora is a convenient alternative.

6. Conclusion

From the investigation, Thai zero anaphor which occurs as the SUBJECT and performs an AGENTITIVE role can refer to first, second, or third person, singular or plural. Moreover, it can refer to all persons according to individual interpretation. In one sentence, there may be more than one zero anaphor and each one may refer to different persons, as in examples 3.1 and 3.2 above. The occurrence of Thai zero anaphor subjects is abundant and they can refer to more antecedents than their English counterparts. The hypothesis of the study is evidently proved true.

We can compare the characteristics of Thai and English zero anaphor SUBJECTS as follows:

<table>
<thead>
<tr>
<th>Zero Anaphor Subject</th>
<th>1st/2nd/3rd person</th>
<th>1st + 2nd +3rd person</th>
<th>one or two in one sentence</th>
<th>more than two in one sentence</th>
<th>same antecedent</th>
<th>different antecedent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thai</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>English</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
</tbody>
</table>

* There are some occurrences but very rare. English uses other grammatical devices, for example, a noun phrase from infinitive or gerund in the place of zero anaphora.

7. Discussion

It is clearly seen that zero anaphor SUBJECTS occur abundantly and freely in Thai. They can perform more functions in the sentence and express more meanings than in English. These different characteristics may cause problems to Thai students. When they speak or write English, the interference from Thai may cause them confusion then they inclined to drop their SUBJECTS and thus commit grammatical errors. In order to raise the students’ awareness, and to help them eliminate these errors, English teachers may prepare supplementary and remedial exercises based on these different characteristics between Thai and English mentioned in this study.

Foreigners who study Thai may apply the results of this study when they are confronted with ambiguous interpretations of zero anaphora. Considering the frequent
occurrence of third-person and all-person antecedents—both types constitute about 84% of the entire occurrence of zero anaphor subjects—they tend to be a refuge for non-Thai natives who are not acquainted with this typical characteristics of the Thai language.

The researcher would also like to suggest for further studies about interpretation of zero anaphora in other Southeast Asian languages that would enrich the global knowledge of linguists and language students.

References
Thai zero anaphor subjects


2004. Matichon, 29th April, 4th May, 3rd June, 16th June.
Art and Culture Magazine. Vol. 20 no. 3
Art and Culture Magazine. Vol. 24 no. 1
2003. Tourism Authority of Thailand Magazine. Vol. 44 December
SHAPE-BASED CLASSIFIERS IN HAKKA AND COMPARISON AMONG THREE DIALECTS

Chiachun Lee
National Tsing Hua University,
Nan-Jeon Institute
chiachun@mail.njtc.edu.tw

Abstract
In the present paper, firstly, we are going to pin down the features of each shape-based classifier in Hakka; secondly, the corresponding classifiers in Taiwanese and Taiwan Mandarin will be compared. The categorization models used are prototype theory, interaction relation, eight categories of classification and dimension saliency. In Hakka, there are about six classifiers relating to shape as their perceptual basis: thiao55, ki31, tsong31, tsak2, te55 and liap4. thiao55 classifies those objects with long shape, in other words, one-dimensional salient objects. The entities take ki31 as classifier have the following features: one-dimensional salient, hard or rigid, the feature of root or sticklike part of the entity when it is used. The first kind of entities taking classifier tsong31 are two-dimensional salient and, on the other hand, are characterized with the function of placement, that means a flat surface to place something. The second type are those which are made from paper. tsak2 categorizes hollow objects which have function of placement and passing through. “box”, “tire”, ”cup” and “teapot” are able to contain things. Moreover, the classifier tsak2 is on its way becoming a general classifier. These entities taking te55 as classifier are featured with a flat surface. The kind of spherical entities are classified by liap4. As far as the comparison of classifier denoting long-shaped entities among these three dialects is concerned, the feature one-dimensional salient is essential feature in all three dialects. Flexible and rigid/hard are two major distinctive features in all three dialects. The feature root, in our opinions, plays a minor but important roll in Taiwanese and especially in Hakka. As for the entities with saliency two dimensions, two-dimensional salient and paper material are two essential features in all three dialects. Spreading or pulling out to use is a feature to be categorized by zhang in Mandarin and tsong31 in Hakka but that’s not found in Taiwanese. It is found that in every set of classifiers the essential features are almost the same. Their differences lie in distinctive features. The paper is organized as following: section 1 is literature review; in section 2 we investigate shape-based classifiers in Hakka in terms of distinctive feature; section 3 the comparison among three dialects is made; section 4 is conclusion.

1.0 Introduction
According to Lakoff(1987), “to understand how human beings categorize in general, one must at least understand human categorization in the special case of natural language.” The choice of classifier in Chinese is anything but an accident. Through the understanding of classifier, we can get a clear picture of how we categorize nouns. There have been a lot of papers concerning about Chinese classifiers. The present paper

1 Please refer to references.
Chiachun Lee

focuses on shape-based classifiers in Hakka. The features of each shape-based classifier are going to be figured out. We adopt the categorization model used in Wu (1996), prototype theory, Denny’s interaction relation, eight categories of classification proposed by Allan (1977), Tversky and Hemenway (1984) and Pinker (1989)’s dimension saliency. In the first section, these categorization models will be reviewed. In the second section, we are going to investigate semantic features of each shape-classifier. In the third section, the comparison of shape-based classifiers in Taiwan Mandarin, Taiwanese and Hakka is made to have a better understanding.

1.1 Prototype theory
Certain members of a category are more representative than the other members of the category. The most representative members of a category are called “prototypical” members. A noun is classified by comparing with the prototypical members.

1.2 Interaction relation
Denny (1976) suggests that “cross-linguistically, classifiers fall into three basic semantic types, all having to do with human interactions: physical interaction such as handling, functional interaction such as using an object as a vehicle, and social interaction such as interacting appropriately with a human compared to an animal, or a high status person compared to a low status one.”

1.3 Eight categories of classification
Based on the seven categories proposed by Allan(1977), Tversky and Hemenway(1984) claims eight categories of classification:

i. material: the essence of the entities. There are three subcategories: animacy, abstract and verbal nouns, inanimacy.

ii. shape: dimensional categories (saliently one-dimensional, two-dimensional, and three-dimensional) and non-dimensional categories (prominent curved exterior).

iii. consistency: flexible, hard or rigid, and non-discrete.

iv. size: big and small.

v. location: the location of a noun.

vi. arrangement: arrangement identify three kinds of objects: I. An object or objects in some specific and non-inherent configuration. II. an object or set of objects in a specific position. III. objects in some kind of specific non-inherent distribution.

vii. Quanta: Quantity is the basis of categorization.

viii. Attributes of parts: the categorization is based on one part of the whole body.
Shape-based classifiers in Hakka

1.4 Dimensional saliency
According to Wu(1996), dimensional saliency could be expressed by $x$, $y$, $z$, three dimensions of an object. ($x>y>z$).

(i) Typical one-dimensional salient, the proportion of $y/x$ is near 0, and that of $z/y$ is near 1.
(ii) Typical two-dimensional salient, the proportion of $z/y$ is near 0, and that of $y/x$ is near 1.
(iii) None salient dimensional, the proportion of $z/y$ is near 1, and that of $y/x$ is near 1.

2.0 Shape-based classifiers in Hakka
There are about six classifiers relating to shape as their perceptual basis: thiao55, kt31, tsong31, tsak2, te55 and liap4. We are going to discuss these classifiers respectively.

2.1 Classifier thiao55
thiao55 classifies those objects with long shape, in other words, one-dimensional salient objects.

(1) objects with feature long

<table>
<thead>
<tr>
<th>String</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>it2 thiao55</td>
<td></td>
</tr>
<tr>
<td>sui22 kong14</td>
<td>a conduit</td>
</tr>
<tr>
<td>phi55 tai22</td>
<td>a leather belt</td>
</tr>
<tr>
<td>so55 e55</td>
<td>a rope</td>
</tr>
<tr>
<td>giu55 thiao55</td>
<td>a fritter of twisted dough</td>
</tr>
<tr>
<td>thu214 si55 mian31 bao55</td>
<td>a loaf of toast</td>
</tr>
<tr>
<td>gieNn31 tshong55</td>
<td>a Chinese sausage</td>
</tr>
<tr>
<td>ngaN55 kao31</td>
<td>a tube of tooth paste</td>
</tr>
<tr>
<td>ne11ku55tai31</td>
<td>a necktie</td>
</tr>
<tr>
<td>bong55 kua31 e55</td>
<td>a cucumber</td>
</tr>
</tbody>
</table>

The long shape characterizes those objects in (1). The proportion of $y/x$ of those objects is near 0, and that of $z/y$ is near 1. This belongs to the type of one-dimensional salient. Those items in (2) are also featured with long shape. And the long feature is the most prominent in shape. Moreover, one can not tell its length without professional measure methods.

(2) objects with feature long

<table>
<thead>
<tr>
<th>String</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>it2 thiao55</td>
<td></td>
</tr>
<tr>
<td>sien31</td>
<td>a line of thread</td>
</tr>
<tr>
<td>lu22</td>
<td>a road</td>
</tr>
<tr>
<td>kai31 lu22</td>
<td>a street</td>
</tr>
<tr>
<td>ho55 pa22</td>
<td>a river</td>
</tr>
</tbody>
</table>

The use of thiao55 could be extended to abstract items, such as in (3). According to Liu(1965), the abstract use originated from that fact that ‘matters’ are recorded on the

---

2 The Hakka we investigate here is Tongluo hakka dialect. The informant is 55 years old, female, and she is a teacher in junior high school in Tongluo.
books which is made of bamboo or wood, the shape of bamboo or wood is long and narrow. This kind of abstract use is based on the shape of concrete object: long and narrow bamboo or wood. In this Hakka dialect, *thiao55* follow the abstract use.

(3) abstract items

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>it2 thiao55</em></td>
<td><em>sin31 bun55</em> news</td>
</tr>
<tr>
<td></td>
<td><em>se22 miaNn22</em> life</td>
</tr>
<tr>
<td></td>
<td><em>khui31 thing22</em> regulation</td>
</tr>
</tbody>
</table>

In our data, *thiao55* also classifies animal items, in (4), that are characterized by long shape. In Tu(1998)'s data, fish is categorized by the classifier *miN55/muiN55*. The tail part of fish is adopted to represent the whole body. Moreover, in Tu's data, “cow” could be classified by *thiao55* in few dialects. That may be explained as that the trunk part of cow is the most salient and important. And the shape of the trunk part is, mostly, long and narrow. However, it is very interesting to find the data that *thiao55* classified animals.

(4) animals

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>it thiao/miN/mui</em></td>
<td><em>ng55 e55</em> fish</td>
</tr>
<tr>
<td></td>
<td><em>sa55 ko31</em> a snake</td>
</tr>
<tr>
<td></td>
<td><em>mo22 mo22 tshuNn55</em> a caterpillar</td>
</tr>
</tbody>
</table>

There is one item that we have to be discussed here, “a rubber band”. When “a rubber band” is not used, it is loose and a circle. When it is used, it is tense and like a rope. “A rubber band” is classified in the form of being used, in other words, its functional interaction with people play a decisive role in categorization.

2.2 classifier *ki31*

Some of the objects with one-dimensional saliency do not take classifier *thiao55*, instead they take *ki31* as their classifier.

(5)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>it2 ki31</em></td>
<td><em>fo55 tsai55</em> a match</td>
</tr>
<tr>
<td></td>
<td><em>bit2</em> a pen</td>
</tr>
<tr>
<td></td>
<td><em>kun22 ne55</em> a stick</td>
</tr>
<tr>
<td></td>
<td><em>dang22 e55</em> a nail</td>
</tr>
<tr>
<td></td>
<td><em>hiong31</em> a stick of incense</td>
</tr>
<tr>
<td></td>
<td><em>tsu22</em> a bamboo stick</td>
</tr>
<tr>
<td></td>
<td><em>thien22 fo22 tun31</em> an electric wire rod</td>
</tr>
<tr>
<td></td>
<td><em>lap2 tsut4</em> a candle</td>
</tr>
<tr>
<td></td>
<td><em>gien55</em> a cigarette</td>
</tr>
<tr>
<td></td>
<td><em>kam55 tsa22</em> a sugarcane</td>
</tr>
<tr>
<td></td>
<td><em>miaNn55 phai55</em> a gambling fortune number</td>
</tr>
<tr>
<td></td>
<td><em>tshiam31</em> a fortune label</td>
</tr>
</tbody>
</table>

---

3 In our data, *li55 iu55* (a reason) is classified by *tsak2* not by *thiao55*.

4 These items in (4) are also categorized by *miN/muiN55* in Tonghuo dialect.
Shape-based classifiers in Hakka

In comparison with these items taking thiao55 as classifiers in (1), these items in (5) are the same with those items in the part of one-dimensional saliency. But they are different in the categorization of consistency. Those taking thiao55 are more flexible, while those taking ki31 are harder and more rigid.

In Taiwan Mandarin, the items with the feature root are categorized by gen. In Hakka dialect, there is not gen classifier. Those items with the feature root take ki31 as classifier.

(6) objects with the feature root

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>it2 ki31</td>
<td>moN55 hair/fur</td>
</tr>
<tr>
<td></td>
<td>thieu55 naN55 moN31 hair</td>
</tr>
<tr>
<td></td>
<td>tsho13 a blade of grass</td>
</tr>
<tr>
<td></td>
<td>tshuNn31 a green onion</td>
</tr>
<tr>
<td></td>
<td>ngaN55 tshi13 a tooth</td>
</tr>
<tr>
<td></td>
<td>su13 a hand</td>
</tr>
<tr>
<td></td>
<td>su55 tsi22 thieu55 a finger</td>
</tr>
<tr>
<td></td>
<td>kiot4 a foot</td>
</tr>
</tbody>
</table>

The original meaning of ki31 is “branch”. The original meaning of thiao55 is “twig or the end part of the branch”. Branches develop from the trunks, twigs develop from branches. If gen, the classifier with root meaning, does not exist, ki31 is a better choice to cover it because ki31 is nearer to the root than thiao55.

Ki31 also categorizes tools. These tools in (7) functionally interacts people with the long, rigid, part as Wu(1999) observes “the sticklike part that is grasped when using the entity, and the classifier chosen is based on a perception of the shape of that part of the entity.”

(7) tools

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>it2 ki31</td>
<td>fu22 thieu55 an ax</td>
</tr>
<tr>
<td></td>
<td>tsa31 e55 an umbrella</td>
</tr>
<tr>
<td></td>
<td>tsien22 to55 a pair of scissors</td>
</tr>
<tr>
<td></td>
<td>su22 tien55 ne55 a flashlight</td>
</tr>
<tr>
<td></td>
<td>so22 ba13 a broom</td>
</tr>
<tr>
<td></td>
<td>kiam13 a sword</td>
</tr>
<tr>
<td></td>
<td>tshun22 e55 a gun</td>
</tr>
<tr>
<td></td>
<td>so22 si55 a key</td>
</tr>
</tbody>
</table>

The entities take ki31 as classifier are characterized by the following features: one-dimensional saliency, hard or rigid, the feature of root or sticklike part of the entity when it is used.

2.3 classifier tsong31

The first kind of entities taking classifier tsong31 are those with two-dimensional salient feature and have the function of placement, that means a flat surface to place something.

(8) objects with feature flat surface.
Another kind of entities are those which should be spread or pulled out to be used.

(9) objects with feature spreading or pulling out to be used.

\[
\begin{array}{ll}
\text{it2 tsong31} & \text{tsoiu13} \\
\text{bong13 e55} & \text{a mouth} \\
\end{array}
\]

It is not accepted to take “ki” as the classifier of “mouth”.

The second type are those which are made from paper. According to Wu(1996), these entities take tsong31 as classifier by analogy with “a piece of paper”. In Hakka, the same analogy also exists.

(10) objects made from paper

\[
\begin{array}{ll}
\text{it2 tsong31} & \text{tsha31 tan22} \\
\text{bui22 sen31 tsi13} & \text{a piece of tissue paper} \\
\text{tsi31 piao22} & \text{a check} \\
\end{array}
\]

2.4 classifier tsak2

tsak2 as a shape-based classifier is very particular. Originally, tsak2 is a classifier for birds and then extends to classify animals.

(11) animals

\[
\begin{array}{ll}
\text{it2 tsak2} & \text{kiou22 e55} \\
\text{tsu31 e55} & \text{a dog} \\
\text{kai31 e55} & \text{a chicken} \\
\text{ap2 e55} & \text{a duck} \\
\text{san22 kiou31 tai22} & \text{a lizard} \\
\text{thi22 fu13} & \text{a leech} \\
\text{bu13 gin55} & \text{a fly} \\
\end{array}
\]

In Hakka, tsak2 categorizes hollow objects that have function of placement and passing through. “box”, “tire”, ”cup” and “teapot” are able to contain things. “head”, which also takes tsak2 as a classifier, is also able to contain things, too. The eyes, which are able to see things, the nose, which can breathe in and out air, and the mouth, which are able to eat food, makes “head” function as a container.  

---

5 Lai55(a plough), which should be pulled out to be used, is not classified by tsong31, but by tsak2 in Tongluo dialect.

6 The development from classifying animals to hollow objects is very interesting. Here we do not tend to find out the path of this development.
Shape-based classifiers in Hakka

(10) hollow objects

\[
\begin{array}{ll}
it2 & tsak2 \\
& siong31 nge55 \quad \text{a box} \\
& len55 e55 \quad \text{a tire} \\
& su22 tsi22 lap4 \quad \text{a ring} \\
& pui22 e55 \quad \text{a cup} \\
& tscha55 kong31 \quad \text{a teapot} \\
& khuNng22 \quad \text{a hole} \\
& si22 \quad \text{a word} \\
& sin22 tai13 e55 \quad \text{an envelope} \\
& thieu55 naN55 \quad \text{a head} \\
\end{array}
\]

According to Liu(1965), the general use of tsak2 as a classifier was rarely found by Nanbei Dynasty. After Tang Dynasty, the general use of tsak2 started to spread, such as “it zhi chuan一隻船(a boat), yi tsi it zhi椅子一隻(a chair)”. But in other dialects, we could not find any general use of this classifier. The classifier tsak2 is on its way becoming a general classifier in Hakka dialect.\(^7\)

2.5 classifier te55
These entities taking te55 as classifier is featured with a flat surface. And the thickness is a decisive factor in this class of categorization because a very thin entity with a flat surface is not categorized by te55. It is suggested that the use of phian31, in (12), is affected by Mandarin. Our informant uses se22te55 (small one), but not phian31 when she heard “a piece of potato chips”. After reminding her that the entity is very thin, then she responds with phian31.

(11)

\[
\begin{array}{ll}
it2 & te55 \\
& thieu22 fu22 \quad \text{doufu} \\
& tscha55 ku31 \quad \text{a bar of soap} \\
& thian55 \quad \text{farmland} \\
& phian55 ne55 \quad \text{a plate} \\
& tson22 ne55 \quad \text{brick} \\
\end{array}
\]

(12)

\[
\begin{array}{ll}
it2 & phian31 \\
& Si31 ku31 \quad \text{a piece of watermelon} \\
& tsoi22 sun55 \quad \text{a piece of lip} \\
\end{array}
\]

2.7 classifier liap4
The kind of spherical entities are classified by liap4. The volume of entity is not so important in choosing the classifier. The big spherical entity is classified by tai22 (large) liap4 and the small one is classified by se22 (small) liap4.

---

\(^7\) Tsak2 seems to be able to replace other classifier in Hakka, that’s to say, tsak2 tends to be a general classifier.
(13) It2 liap4 miN13 a grain of rice
thi22 thieu22 a peanut
long13 an egg
san31 a mountain
liu55 lien55 a durian
po31 li55 tshoiu22 A cabbage

3.0 Comparison
In this section we try to compare shape classifiers in Taiwan Mandarin, Taiwanese and Taiwan Hakka in terms of perceptual basis and semantic field.

3.1 Zhi,thiao,gen in Taiwan Mandarin, ki,tiau in Taiwanese and thiao55,ki31 in Hakka

<table>
<thead>
<tr>
<th>features</th>
<th>Taiwan Mandarin</th>
<th>Taiwanese</th>
<th>Hakka</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-dimensional saliency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flexible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigid/hard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kí</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>tiau</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Gen</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

According to Lien and Wang (1999), ki in Taiwanese is almost the union of zhi and gen in Taiwan Mandarin. Wu (1999) also observes that ki in Taiwanese overlaps with the classifier gen in Taiwan Mandarin. In Hakka, the similar distribution of ki31 is found. Ki31 covers the scope of the entities with “root”. The feature one-dimensional saliency is essential feature in all three dialects. Flexible and rigid/hard are two major distinctive features in all three dialects. The feature root, in our opinions, plays a minor but important roll in Taiwanese and especially in Hakka.

---

8 These features are concluded from Wu(1996), Wu(1999) and Lien & Wang(1999).
3.2 zhang in Mandarin, tiuNn in Taiwanese and tsong31 in Hakka.

<table>
<thead>
<tr>
<th>features</th>
<th>Mandarin</th>
<th>Taiwanese</th>
<th>Hakka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-dimensional saliency (flat surface)</td>
<td>Spreading or pulling out to use</td>
<td>Paper material</td>
<td>Two-dimensional saliency (flat surface)</td>
</tr>
<tr>
<td>Zhang</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table II

In all three dialects, entities with saliency two dimensions, in other words, a flat surface for placement, and paper material take zhang/tiuNn/tsong as classifier. But in Taiwanese, the entity that should be spread or pulled out to use is not categorized by tiuNn.9

**Two-dimensional salient** and **paper material** are two essential features in all three dialects. **Spreading or pulling out to use** is a feature to be categorized by zhang in Taiwan Mandarin and tsong31 in Hakka but that’s not found in Taiwanese.

3.3. khe, li in Mandarin, liap in Taiwanese and liap4 in Hakka

<table>
<thead>
<tr>
<th>features</th>
<th>Mandarin</th>
<th>Taiwanese</th>
<th>Hakka</th>
</tr>
</thead>
<tbody>
<tr>
<td>spherical</td>
<td>big</td>
<td>small</td>
<td>spherical</td>
</tr>
<tr>
<td>khe</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>li</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table III

In Mandarin, ke and li are different in the size of the entity. If it is larger, ke is preferred, but if it is smaller, li is preferred. The size is not the dominant factor to choose khe or li. However, both in Taiwanese and Hakka there are only one classifier liap to cover the entities with spherical feature.

3.4 phian/khuai in Mandarin, phiN/te in Taiwanese and phian/te/khuai in Hakka

<table>
<thead>
<tr>
<th>features</th>
<th>Mandarin</th>
<th>Taiwanese</th>
<th>Hakka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness</td>
<td>flexibility</td>
<td>flat surface</td>
<td>restricted</td>
</tr>
<tr>
<td>phian</td>
<td>--</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>khuai</td>
<td>+</td>
<td>--</td>
<td>+</td>
</tr>
</tbody>
</table>

Table IV

9 The entities that should be spread or pulled out to use take different classifiers in Taiwanese. For example, bang nga (net) takes niaN, tshui (mouth) takes ki.
The features thickness and flexibility play important parts in choosing the use of *phian* or *khuai* in Mandarin. But in Taiwanese, the scopes of these two classifiers overlap. The entities with the feature thinness and flexibility could take *te/phiN* as classifier. Take two entities for examples: a piece of chewing gum and a piece of potato chips. On the other hand, in Hakka, there is only one classifier to categorize these kinds of objects. In Taiwanese and Hakka, *te* are used frequently than *phiN* or *phian*. It is more acceptable to say that the semantic scope of *te* is much more larger than that of *phiN* or *phian*. **Thickness** and **flexibility** function as two distinctive features to distinguish *phian* and *khuai* in Mandarin. **Thickness and flat surface** are two major components to be classified by *te* in Hakka and Taiwanese. In Taiwanese, *tua phiaNn e te* (a large area of land) and in Hakka *thai phian* is used to modified grassland.

3.5 *ge* in Mandarin, *e* in Taiwanese and *tsak* in Hakka
According to Wu (1996), *ge* has its own meaning, classifying objects with hollow feature, and also functions as a general classifier to replace another classifier in conversation. *Tsak2* in Hakka has its own meaning and tend to be a general classifier. But in Taiwanese, *e10*, without its own meaning, mainly functions as a general classifier to replace another classifier which has been used in earlier conversation. There are some reasons that each dialect tends to develop its own general classifier. Firstly, to avoid repetition, especially found in conversation. Secondly, people, especially foreigners and children when they are learning Chinese, are not sure to use which classifier. They will choose the “safest” one, that means, the one used the most frequently. Thus, it could be clearly found that the general classifier in these three dialects tend to be the one used the most frequently. Thirdly, it is for the purpose of communication.

4.0 Conclusion
From section 2, we have a clear picture of feature analysis of these shape-based classifiers in Hakka. In section 3, these shape-based classifiers are compared. It is found that in every set of classifiers the essential features are almost the same. Their differences lie in distinctive features. A feature maybe carries the function to distinguish two subset classifiers in one dialect; however, this feature does not have the same function in another dialect. These features are constructed differently in hierarchy. Their highest features are the essential ones, the lower ones are less weak in choosing which classifier to use. And it is very interesting to find out that each dialect has its own general classifier even though some general classifiers are on their ways to be a general classifiers. In order to have a better and insightful understanding of classifier in every dialect, the further study is demanded.

---

10 From Chen(1958), there is a general classifier in Southern Min, *ge24*. Its distribution is similar to that of *ge* in Mandarin. We are not sure which dialect of Southern Min is recorded by Chen. But it is believed that there is a classifier tend to be a general classifier in each dialect in Chinese.

11 The ideas are the result of discussion with Professor Tsao in class.
Shape-based classifiers in Hakka

References


Tu, Chunjin. 1998. *Taiwan Zhongbudichu Kejia Fanyian Cihui Duizhaobio (The lexical correspondences of Hakka dialects in the middle part of Taiwan)*, grant from National cultural art foundation.

THE LOGICAL CONSTRUCTIONS OF CONDITIONAL SENTENCES AND THE FUNCTIONS OF CONDITIONAL EXPRESSIONS: A CONTRASTIVE STUDY OF JAPANESE AND THAI

Saowaree Nakagawa
NIT International Center, Nagoya Institute of Technology
nakagawa.saowaree@nitecb.ac.jp, noi@cty-net.ne.jp

1. Introduction

In a previous study I proposed another category of modality which I called Cognitive Modality: Causal, Temporal and Hypothetical, in order to explain sufficiently the meaning of conditional expressions in Japanese and Thai. In this paper, I focus on the logical constructions of conditional sentences and the functions of conditional expressions used in these sentences in both languages; namely, thãa, hàak, mûaa and phã in Thai and ba, to, tara and nara in Japanese respectively. The term logical construction used in this paper means the mental attitude of the speaker at the moment of utterance (I use P to refer to the antecedent and Q to refer to the consequence of a conditional sentence).

Look at the following sentences:

(1) Jikan ga at tara, sentorea kuukoo e kengaku ni ikitai desu ne.
(2) Eki ni tsui tara denwa o shite kudasai.
(3) Kinoo Tookyoo e it tara Yamada san to atta.

If we translate (1) into Thai, it will be as (1)’

(1)’ thãa mii wee-laa yàak pay thîaaw sà-nãam-bin sen-thrã-ææ nà.

Any native speaker of Thai will know that (1)’ is a hypothetical sentence.

When one translates sentence (2) into Thai, one would add lææw (which expresses the meaning of completion ) after thãa since sentence (2) holds the meaning of temporal supposition.

(2)’ thãa thûun sà-thãa-nii lææw chûaay thoo-râ-sàp maa nà.

In the case of sentence (2)’, the word thãa does not convey a hypothetical meaning anymore. Therefore thãa can be omitted leaving lææw alone to express temporal relation of P and Q. In sentence (3), the word tara in Japanese is used. However, this sentence will be expressed in Thai by using the word tsor, phã or mûaa as shown below:
One should notice that in (3)’, both a and b express the time of the occurrences in Q. However, both とonn and みああ do not hold the meaning of being surprised as in タラ in Japanese. Thus, the adverb どうよう がん (unexpectedly) is required.

Let’s look at sentence (4) below:

(4) Mado o aketara/to, yuki ga futte ita.
When I opened the window, I found that it had been snowing.

In the case of sentence (4), the action of opening the window in P and the discovery in Q occur in an instant. Thus, ところ is used instead of とonn and みああ. Panthumetha (1984) explains the meaning of みああ that it denotes the occurrences in P and Q taking place at the same time. However, this does not imply that the two occurrences happen in an instant. (Nakagawa p.114)

(4)’ ところ な-たあSanז ごู่่ phosph wåa ひ-マトòk.
When I opened the window, I found that it had been snowing.

Look at sentence (5) and sentence (6) of Japanese and the translations in Thai below:

(5) Haru ga kureba/to, hana ga saku.
When spring comes, the cherry trees blossom.
(5)’ みああ/ところ รัู่ญรูู้ง ถ่ัญดููย์ ดํัญ-าญ์ โกํ์ บาน.

(6) Taro wa okane ga aru to, ryokoo o suru.
Whenever Taro has money, he goes traveling.
(6)’ ところ ทํา-รููโหม ngữ แกํ ผููญ กํ์ ผูาย ทุูาว.

Sentence (5) and sentence (6) denote a causal meaning, that is, whenever P takes place Q will occur.

2. Cognitive Modality
Most grammarians explain the modality of conditional expressions in Japanese on the basis of objectivity and subjectivity. However, as illustrated in the sentences above, it is obvious that if one judges from the standpoint of objectivity and subjectivity this would not adequately explain Japanese conditionals and this applies to Thai conditional expressions as well. In the previous study, I proposed that in order to explain the meaning of conditional expressions in Japanese and Thai, one should know the speaker’s knowledge of what is true or not true in the domains of realis and irrealis. As Palmer (1986) points out, in conditionals, there exist subcategories of the speaker’s mental attitude. Thus, in the previous study I proposed another scale of mental attitude of the speaker along the scale of subjectivity and objectivity and this is the scale of what I call Cognitive Modality. Thus, the term modality used in the present study means the speaker’s knowledge of P which can be divided into three categories: causal, temporal
The logical constructions of conditional sentences

and hypothetical. In the previous study, I proposed the scale of cognitive modality which is illustrated in Fig.1. It should be mentioned here that unlike Akatsuka (1983), in this study both the domains of realis and irrealis are in the subjective world and they are on a continuum. The definitions of the three categories of modality are as follows:

a) **causal modality**: At the moment of utterance, the speaker believes that P is a fact and whenever P occurs Q would occur repeatedly; that is, P is the cause of the occurrence in Q as illustrated in sentences (5) and (6) above.

b) **temporal modality**: At the moment of utterance, the speaker believes that P is true or P will be realized in the future time. This is illustrated in sentence (2) and sentence (3), see p.41.

c) **hypothetical modality**: At the moment of utterance, the speaker knows that it is possible that P would realize or the speaker knows that it is impossible that P would be realized. Sentence (1) belongs to this category.

**Fig.1: Japanese and Thai cognitive modality and the epistemic scale**

<table>
<thead>
<tr>
<th>Tim / Space</th>
<th>Actual World</th>
<th>Subjective World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past</td>
<td>Actual</td>
<td>Realis</td>
</tr>
<tr>
<td>Future</td>
<td></td>
<td>Irrealis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linguistic World</th>
</tr>
</thead>
<tbody>
<tr>
<td>realized, unexpected, temporal</td>
</tr>
<tr>
<td><strong>Japanese</strong></td>
</tr>
<tr>
<td><strong>Thai</strong></td>
</tr>
</tbody>
</table>

3. Prototypes of conditional expressions

It can be seen that many of the conditional expressions in both Japanese and Thai can be replaced by others. In other words, many conditional expressions have a usage in their own domain and at the same time they can be used in other domains. This notion of expansion has been introduced by Masuoka (1993). In the previous study, I adopted the notion of characteristic expansion in explaining the meaning and usage of these conditional expressions and I divided Cognitive Modality into 3 prototypes, namely “Causal Prototype”, “Temporal Prototype” and “Hypothetical Prototype” as shown in Fig.2.
Saowaree Nakagawa

Fig.2: Prototypes of conditional expressions in Japanese and Thai

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Japanese</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal Prototype</td>
<td>ba, to</td>
<td>mûaa, phôô, thâa, hàak</td>
</tr>
<tr>
<td>Temporal Prototype</td>
<td>tara, to</td>
<td>mûaa, phôô, thâa, hàak</td>
</tr>
<tr>
<td>Hypothetical Prototype</td>
<td>ba, to, tara, nara</td>
<td>thâa, hàak, mûaa</td>
</tr>
</tbody>
</table>

4. Logical Constructions of conditional sentences and the functions of conditional expressions

One can observe that each prototype shown in the table above has various expressions. Causal Prototype, for example, has ba and to in Japanese and mûaa, phôô, thâa, hàak in Thai. One question would arise, what is the difference among them? To answer the question, it is significant to find out what the logical construction the conditional sentence has in using each of these expressions.

a) Causal modality

As explained before, causal meaning denotes that at the moment of utterance, the speaker believes that P is a fact and whenever P occurs Q would occur repeatedly; that is, P is the cause of the occurrence in Q. However, let’s look at the following sentences again:

(5) Haru ga kureba /to, hana ga saku.
    When spring comes, the cherry trees blossom.
(5)’ mûaa /phôô thûû rûû-dûû bây-mâyphî dûûk-mây kûû baan.

(6) Taro wa okane ga aru to, ryokoo o suru.
    Whenever Taro has money, he goes traveling.
(6)’ phôô thà-rôô mîi nûûn kûû pay thûaaw.

One can notice that sentence (5) in both languages indicate general events of which ba or to is used in Japanese and mûaa, phôô is used in Thai. In sentence (6), only to in Japanese and phôô in Thai is used respectively. One can see that in sentence (5), the occurrence in P and the occurrence in Q is related in terms of temporal and it is a general matter while the event in (6) is a specific matter. Look at sentence (7)

(7) pà-kà-tí thûaâ lûûk-khâa mîi nûûy ràán kûû pît rew.
    Usually when there are few customers, the shop will close early.

We can see that thûaâ, which expresses a hypothetical meaning, can be used in a causal relationship between P and Q especially for specific events. In this case, the word pà-kà-tí (usually) is optional. However, when thûaâ is used, it implies the nuance of possibility.

The logical construction of both general and specific events of causal meaning: The speaker believes that when the occurrence in P is realized, the occurrence in Q will take place and this will occur repeatedly.
The logical constructions of conditional sentences

In Japanese, both *ba* and *to* can be used to denote causal meaning. However, *ba* and *to* in Japanese, as Matsushita (1928) explains, have different nuances. That is *ba* expresses the meaning of causal relation of P and Q on logical reason while *to* denotes the meaning of causal relation between P and Q in the sense of actual event. Thus, one can see that in sentence (6) which is a specific event and the speaker focuses on the actual causal relation of P and Q rather than the time reference as in (5), *to* is used. Surprisingly, in Thai, we can also see that when the speaker focuses on the cause-effect, like *to* in Japanese, *phɔɔ*, which has the meaning of cause-effect, is used. Thus, the logical construction of the sentence of causal relation expressed by each expression can be categorized as shown in Fig.3 (t is used for true, f is used for false).

Fig. 3: The logical constructions of causal modality

<table>
<thead>
<tr>
<th>Conditional Sentence</th>
<th>Conditional Expressions</th>
<th>P</th>
<th>Q</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Japanese</strong></td>
<td><em>ba</em></td>
<td>t</td>
<td>t</td>
<td>logical causal relation</td>
</tr>
<tr>
<td></td>
<td><em>to</em></td>
<td>t</td>
<td>t</td>
<td>actual causal relation</td>
</tr>
<tr>
<td><strong>Thai</strong></td>
<td><em>mùaaa</em></td>
<td>t</td>
<td>t</td>
<td>temporal</td>
</tr>
<tr>
<td></td>
<td><em>phɔɔ</em></td>
<td>t</td>
<td>t</td>
<td>actual causal relation</td>
</tr>
<tr>
<td></td>
<td><em>pàkà-ti... thâa/hàak</em></td>
<td>t</td>
<td>t</td>
<td>possibility</td>
</tr>
</tbody>
</table>

Note that both *to* in Japanese and *phɔɔ* in Thai are used in statements describing events, in other words, the two words are used for sentences describing propositions in P and Q. Thus, expressions such as demand, request and so forth can not be used with *to* and *phɔɔ*.

b) Temporal modality
Temporal modality means that at the moment of utterance, the speaker believes that P is true or P will be realized. This is illustrated in sentence (2) and sentence (3).

(2) Eki ni tsui *tara* denwa o shite kudasai.
When you arrive at the station, give me a call.
(2)’ *thâa* thúnj sá-thâa-nii laëæw chûaay thoo-rá-sàp maa ná.

(3) Kinoo Toorkyoo e it *tara* Yamada san to atta.
Yesterday when I went to Tokyo, I happened to meet Mr. Yamada.
(3)’ *mùaaa* pay too-kiaw chân phóp khun yaa-maa-dà dooy bãñ-çon.

(4) Mado o ake *tara/to*, yuki ga futte ita.
When I opened the window, I found that it had been snowing.
(4)’ *phɔɔ* pèòt nà-tàaj kôö phóp wâa hû-má tôk.
When I opened the window, I found that it had been snowing.

(8) Taro wa heya ni hairu to denki o tsuketa.
When Taro went into the room, he turned on the light.
(8)’ *phɔɔ* thaa-roo khâu hòeñ kôö pèòt fay.
In Japanese, *tara* has been treated as a conditional expression expressing with a high degree of subjectivity. However, Akatsuka (1983) cites that *tara* also expresses temporal meaning. Sentence (2) is one example of temporal meaning. In this sentence, the speaker believes that the occurrence in P will be realized in the near future and therefore the speaker expresses a request in Q. Since *tara* implies a high degree of subjectivity, it can be used with sentences expressing commands, requests and so forth.

In the Thai language *thāa*, which is usually used to express hypothetical meaning, can be used to express the meaning of temporal when used together with *lō难以* (which denotes the meaning of completion). When denoting temporal meaning, *thāa* is optional.

Sentence (3) indicates that at the moment of utterance, the speaker knows P is true and it occurred in the past which was an unexpected event. However, since such a sentence expresses an occurrence in the past which can be replaced by *toki* (when), it is included in the temporal category in this study. In Japanese, *tara* and *to* can be used with no significant difference. In Thai, *muāa* and *phos* are used with different meanings.

In the case of sentence (4), the action of opening the window in P and the discovery in Q occur in an instant. Thus, *phos* is used instead of *toon* and *muāa*. As mentioned previously, Panthumetha (1984) explains the meaning of *muāa* that it denotes the occurrences in P and Q taking place at the same time. However, this does not imply that the two occurrences happen in an instant. The meaning of sentence (3) is ambiguous; it could be that meeting Mr. Yamada took place right after having arrived in Tokyo or took place any time during my stay in Tokyo. Therefore *muāa* and can be used but not *phos* since *phos* denotes merely an instant occurrences of P and Q. In sentence (8), the speaker expresses the occurrences in P and Q which occurred in an instant; therefore, *to* and *phos* are used in Japanese and Thai respectively. See Fig. 4

**Fig. 4: The logical constructions of temporal modality**

<table>
<thead>
<tr>
<th>Conditional Sentence</th>
<th>Conditional Expressions</th>
<th>P</th>
<th>Q</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Japanese</strong></td>
<td></td>
<td>t</td>
<td>t</td>
<td>temporal</td>
</tr>
<tr>
<td><em>Tara</em></td>
<td></td>
<td>t</td>
<td>t</td>
<td>past event : discovery</td>
</tr>
<tr>
<td><em>Tara, to</em></td>
<td></td>
<td>t</td>
<td>t</td>
<td>past event: instant</td>
</tr>
<tr>
<td><em>to</em></td>
<td></td>
<td>t</td>
<td>t</td>
<td>temporal</td>
</tr>
<tr>
<td><strong>Thai</strong></td>
<td></td>
<td>t</td>
<td>t</td>
<td>past event: discovery</td>
</tr>
<tr>
<td><em>thāa</em></td>
<td><em>hāak...lō难以</em></td>
<td>t</td>
<td>t</td>
<td>-instant</td>
</tr>
<tr>
<td><em>muāa</em></td>
<td><em>toon...dooy baŋ-soon</em></td>
<td>t</td>
<td>t</td>
<td>+instant</td>
</tr>
</tbody>
</table>

46
c) **Hypothetical meaning**

At the moment of utterance, the speaker knows that it is possible that P would be realized or the speaker knows that it is impossible that P would be realized. Look at the following sentences:

(1) Jikan ga at tara/ ba, sentorea kuukoo e kengaku ni ikitai desu ne.  
   If I have time, I would like to go to see Centrair Airport.

(1)’ thàa mii wee-laay yàak pày thàaw sà-nàam-bin sen-thòô-àeëe ná.

(9) Hayaku okinai to maniawanai yo.  
   If you don’t get up early, you won’t be in time.

(9)’ thàa mày ríip tuûun kòô cà mày than ná.

(10) motto benkyoo sureba, seiseki ga agaru deshoo.  
    If you study harder, your grade would probably be better.

(10)’ thàa / hàak / múàaa thòô khà-yànn riaa kwàa níi khàn-àeën khoç cà dìi khùûn.

(11) Yamada san ga dekirunara, watashi mo dekiru.  
    If Mr. Yamada can do it, I can do it, too.

(11)’ thàa / múàaa khun yàa -mà-à da thàm-dày chànn kòô thàm-dày.

(12) Edojidad ni umarete ireba, Tokugawa ieyasu ni acta kamoshirenai.  
    If I were born in the Edo era, I might have been able to see Ieyasu Tokugawa.

(12)’ thàa / hàak chànn kòôt này sà-mày ee-dòÔ chànn âát cà mìi oo-kàat phòp kàp too-kù-ŋàa-wàâ ii-ee-yàa-suí kòô dày.

The logical constructions of sentence(1) and (9)~(12) can be explained as follows: The speaker is not certain that P will be realized, but if it is realized, then Q will occur. In Japanese *ba* is used when the speaker wants to express formality.

In sentence (9), The speaker is not certain that P will be realized, in case it is realized, Q will occur and usually in an undesirable result.

In sentence (10), the speaker utters a sentence expressing logical hypothetical relation between P and Q. Usually, when Q is a desirable result, *ba* will be used in Japanese while *thàa*, *hàak* and *múàaa* can be used freely.

In sentence (11) in both Japanese and Thai, the speaker utters the sentence after having learned the new information from the listener. The speaker expresses the hypothetical attitude towards the information just learned from the listener implying the meaning that “if what has been told is true, then Q will occur”.

Sentence (12), the speaker knows that P is not true and Q would never be realized. The logical construction of the hypothetical modality can be summarized in Fig. 5.
### Conditional Sentence and Expressions

<table>
<thead>
<tr>
<th>Conditional Sentence</th>
<th>Conditional Expressions</th>
<th>P</th>
<th>Q</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ba</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, possible, desirable result, formality</td>
</tr>
<tr>
<td></td>
<td><em>to</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, possible expressing undesirable result</td>
</tr>
<tr>
<td></td>
<td><em>tara</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, possible</td>
</tr>
<tr>
<td></td>
<td><em>nara</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical towards information, possible</td>
</tr>
<tr>
<td></td>
<td><em>ba, to, tara, nara</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, impossible</td>
</tr>
<tr>
<td>Thai</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>tha</em> , <em>hàak</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, possible</td>
</tr>
<tr>
<td></td>
<td><em>hàak</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical possible, expressing formality</td>
</tr>
<tr>
<td></td>
<td><em>tha</em> , <em>hàak</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical, impossible</td>
</tr>
<tr>
<td></td>
<td><em>muàa</em></td>
<td>f</td>
<td>f</td>
<td>hypothetical towards information, possible</td>
</tr>
</tbody>
</table>

### 5. Conclusion

As illustrated in the examples above, it can be concluded that in explaining Japanese and Thai conditional expressions, it is crucial to examine the mental attitude of the speaker at the moment of utterance towards the realization of P in the domain of realis and irrealis. From this, one can know the semantic characteristics of each expression that can be appropriately used for the suitable conditional sentence to fit its logical construction. The findings of the study of the logical constructions of conditional sentences and the functions of the conditional expressions are as follows:

1) In both Japanese and Thai, conditional expressions in the same prototype; namely causal prototype, temporal prototype and hypothetical prototype have different nuances and functions.

2) In Japanese *to* has a wide range of usage while in Thai language *tha* is used widely with different meanings.

3) *tara* in Japanese and *tha* in Thai share the similarity in that they both express hypothetical as well as temporal meanings. Besides, *to* in Japanese is similar to *pha* in Thai in many aspects; they are used to combine two events that occur in an instant and both share the same feature in expressing causal relation of P and Q.

4) *ba* in Japanese is similar to *hàak* in Thai as expressing formality.

5) *nara* and *muàa* express hypothetical meaning towards the information given by the listener of the conversation.

6) *to, tara* in Japanese is used in the past event and it expresses the meaning of
discovery an unexpected event. In Thai, phɔɔ is used with the same meaning. However, unlike to and tara in Japanese, phɔɔ is used in the case of instant occurrences only. When it is a discovery and the two events in P and Q do not occur in an instant, either muąa...dooy ban-üler or tećć...dooy ban-üler is used.

References
----- 1985 “Conditionals and The Epistemic Scale” Language 61(3):625-639
Masuoka, T. 1993. Nihongo no Jookken Hyoogen, Kuroshio Shuppan
------ 1993. Jookken Hyoogen to Bun no Gainen reberu, Kuroshio Syuppan
Matsushita, D. 1928. Kaizen Hyoojun Nihongo Bunpoo, Tokuda Masanobu Hen
Teramura, H. 1979. Bunpoo Kenkyuu — Muudo no Keisiki to Imi—Gaigenteki Hoodoo no Hyoogen, Bungei Gengokenkyuu Gengo hen (4) University of Tsukuba
Thonglaw, K. 1957. Thai Standard Grammar. Ruamsan Publisher.
Abstract
Previous research on the syntactic typology of Vietnamese has led to two controversial findings. One set of studies has claimed that the language is a subject-prominent language (Ly 1948, Chinh & Le 1973, Ban 1987, Thin 2001). The other set states that the basic structure of Vietnamese manifests a topic-comment relation rather than a subject-predicate relation (Thompson 1965, Dyvik 1984, Hao 1991, Anh 2000). This uncertainty between these two schools of thought has resulted in an undeniable inconsistency in the definitions and classification of different types of Vietnamese sentences among Vietnamese and non-Vietnamese scholars.

With the assumption that Vietnamese is a topic-prominent language, the aim of the study is to investigate to what extent the typological differences between the two languages influence the process of translating authentic Vietnamese sentences into English, through a preliminary report on an error analysis of the Vietnamese-English translations of Vietnamese EFL students. The subjects include 95 students from the first, second, third, and fourth years of the Department of English Language and Literature at the University of Social Sciences and Humanities, Ho Chi Minh City, Vietnam. The data will be the translation texts of about 95 students from the same source text. As a translation error may be attributable not only to a lack of linguistic competence but also to a lack of translation competence and a deficiency in comprehension ability of the source text, the study hopes to build up a taxonomy of specifically translation errors, especially in the field of translation into a second language. Hopefully this taxonomy will open up a possible way to prevent future errors from learners in translating the topic-comment structures of Vietnamese, as well as helping teachers in teaching Vietnamese-English Translation.

This study also suggests some practical guidelines in applying the techniques of error analysis into the teaching of Vietnamese-English translations, especially topic-comment structures of Vietnamese.

Introduction:
Although many studies have been carried out in error analysis and contrastive analysis in second language learning, language teaching and materials development, not much research has been done in the Vietnamese context with Vietnamese university students as informants. To illustrate, in a comprehensive bibliography by Spillner in the field of error analysis (1991), of the 2% (108) of studies focusing on translation out of a total of 5,398, none looked at the syntactic errors in Vietnamese-English translation. This study aims to fill this gap in the field of error analysis, especially errors in Vietnamese-English translation, with the hypothesis that the Vietnamese topic-comment structure...
and its empty elements can cause some difficulty in the translation process. It is hoped that the present study will shed light on the most common types of errors by Vietnamese students in translation and that it will have implications for translation pedagogy.

Although these errors cannot be completely considered as translation errors as students do not receive official training in a four-year course specialising in translation, they are errors made during the course of studying translation. Therefore, any problems found from these errors can help the material designers and teachers to choose an appropriate method for their teaching. It is stressed that this study is more to serve the translation and language teaching rather than translation per se and the subjects are also advanced second language learners.

**Literature Review**

**Error Analysis in Translation**

Firstly, even excellent translators make mistakes in translations. Secondly, some errors are almost unavoidable given the fact that translators and interpreters inevitably have vocabulary and knowledge gaps. Neubert & Shreve (1995) describe a translation error in the following statement:

> What rightly appears to be linguistically equivalent may very frequently qualify as ‘translationally’ nonequivalent. And this is so because the complex demands on adequacy in translation involve subject factors and transfer conventions that typically run counter to considerations about ‘surface’ linguistic equivalence (1995:415).

This statement partially describes the complication and difficulty in defining and identifying translation errors. Translation errors are different from errors that would occur in spontaneous native language production. In translation, working with a source text induces errors under the influence of source language morphology, whereas in spontaneous second language production, native morphological system of language learner tends to interfere with knowledge of the second language system. In the case of second language learners, identifying translation errors is tricky as translation errors may be mixed up with linguistic errors.

In the literature of translation training, many studies have been done to find out what types of errors translators/interpreters/student translators often committed in their process of rendering a certain structure from one language into another (Altman 1994, Coskun 1997, Dodds 1999, Gile 1994, Seguinot 1990). These studies are based on the premise that the insight into the act of making error can shed light to the psycholinguistic (mental process) of translators/interpreters/student translators and contribute to the training of translators.

**Topic-Comment Structures in Vietnamese**

Topic and Comment are the two concepts which have been constructed differently by various linguists. Different definitions on these two terms have existed, under a number of different names and guises: presupposition and focus (Chomsky 1971, Jackendoff 1972), theme and rheme (Firbas 1972) topic and comment (Gundel 1974, 1978) open proposition and focus (Ward 1985 and Prince 1986) and ground and focus (Vallduvi 1990). Despite of the difference in the details of these various descriptions, the concept of topic and comment are in general based on the intuitions that utterances we say are ‘about’ something (topic) linking up with information the speaker assumes the hearer is
Errors in the translation of topic-comment structures

aware of, and that utterances contain information the speaker is presenting as new relative to this topic (comment).

The topic of a sentence is basically what the sentence is about. It always appears at the very beginning of a sentence, referring to something that the speaker assumes the listener would have some knowledge. A topic does not equal to a subject of a sentence in that a subject must always have a direct semantic relationship with the verb as the one that performs the action or exists in the state indicated by the Verb, yet a topic need not. It can be followed optionally by a pause in speech or a comma in writing, showing what is being talked about, apart from the rest of the sentence.

Topic is often defined in terms of its linguistic structures, either syntactic or phonetic. It has been defined in terms of linear order – as the first expression of the sentence (e.g. Halliday, 1967), in grammatical terms – as the subject (Gundel 1974) and in intonational terms – as the non-stressed expression (Chomsky 1971). However, the shortcomings of these definitions lie in their inability to answer the question related to the discourse conditions under which a given expression would count as topic.

In their seminar article ‘Subject and Topic: a New Typology of Language’, Li and Thompson (1976) have set up many criteria to distinguish the difference between the notion of subject and that of topic. The main difference, according to them, lies in the idea that while topic is a discourse-related notion, subject is more integrated into the syntax of the sentence. They also argued that topic should be treated as a basic rather than a derived category. The treatment of topic as one basic category will entail the classification of language into two different types: Subject-prominent language and Topic-prominent language.

Using Li and Thompson’s (1976) argument as a theoretical standpoint, Hao (1991), a Vietnamese scholar, examines Vietnamese from a functional approach. According to Hao (1991), the way the Vietnamese express themselves is that ‘when uttering a sentence, the speaker produces a topic and says something about that topic or within the range of that topic’ (1991:79). That is to say, when re-organising the reflected reality, thought divides it into two parts by choosing a point of departure for establishing the relationship between these two. He assumes that the part that is chosen as the point of departure functions as topic and the remainder as comment (1991:33-4). In his opinion, the topic-comment structure in the Vietnamese sentence is a phenomenon which belongs to what he refers to as the ‘logico-discursive domain’.

In support of his claim that Topic-Comment is the dominant structure of Vietnamese, Hao provides two reasons. His first reason is derived from the result of the study by Li & Thompson (1976), who claimed that there are four main types of language: (i) languages that are subject-prominent (e.g., Indo-European, Niger-Congo, Fino-Ugric, etc), (ii) languages that are topic-prominent (e.g., Chinese, Lahu, Lisu etc), (iii) languages that are both subject-prominent and topic-prominent (e.g., Japanese, Korean, etc.), and (iv) languages that are neither subject-prominent nor topic-prominent (e.g., Tagalog, Ilocano etc.). Hao (ibid.) claims that like Chinese, Vietnamese belongs to category (ii). The second reason comes from his reality of using Vietnamese. According to Hao’s calculations, only 30% of Vietnamese sentences are of Subject-Predicate type while about 70% of them are of Topic-Comment type.

However, this study will not use Hao’s definition of topic-comment for data analysis, mainly because his classification is still ambiguous and not facilitating. The present study will utilise the classification that Rosen (1998) has proved in her PhD thesis. According to her, topic-comment structures and empty elements are basic
Pham Phu Quynh Na

constructions in Vietnamese language. There are five types of relations that may exist between a Noun Phrase topic and the comment in Vietnamese.

1. The topic may be understood as filling a gap in the comment.
2. The topic may be co-referential with a noun phrase or pronoun in the comment.
3. The topic may be semantically related to any specific constituent in the comment, its referent is simply what the comment is about.
4. The topic may not be related to any constituents, empty or overt, in the comment.
5. The topic may be a verb phrase, a clause or an embedded topic-comment construction.

Regarding the empty pronoun, there are four types in Vietnamese:

6. Empty pronouns in simple sentences
7. Empty pronouns in embedded clauses
8. Empty Pronouns in consecutive clauses
9. Empty Pronouns with indefinite reference

In the Elicitation Task of this study Sentence 1 has the empty pronoun in embedded clauses as stated in (7). The explanatory sentence at the end of Sentence 1 (inside the bracket), sentence 14, 15, 16 are examples of the empty pronouns in consecutive clauses as stated in (8). Sentence 2 and 13 are the examples of empty pronoun in simple sentences as stated in (6). Sentence 5 is the example of sentence where the topic is semantically related to a specific constituent in the comment, as stated in (3) (the topic is the object of the verb mentioned in the comment). Sentence 6 is the example where the topic is not related to any constituents, empty or overt, in the comment (the topic is the adverb of the comment) as stated in (4). Sentence 7, 8, 9, 10 are elliptical sentences. Sentence 17 consist of two comments, the first comment is an embedded topic-comment structure (leading to the so-called ‘double-subject construction’), the NP topic of which is co-referential with the main topic, as stated in (2). The second comment of sentence 17 is also an embedded topic-comment structure, the topic of which is semantically related to the main topic of the sentence, as stated in (3). Sentence 18 also has two comments: the first comment is an embedded topic-comment structure, the second one is an embedded one but there is a gap within this second comment which the main topic of the sentence fills up as stated in (3) [see Appendix A for numbered sentences in the Test and Appendix B for the analysis of topic-comment constructions and empty pronouns of 18 sentences].

Research Questions

a. Do Vietnamese students have problems in translating Vietnamese sentence types in which the Topic-Comment structure is more prominent than a Subject-Predicate one? What sort of errors do they make in translating this specific type of structure?

---

1 Constructions have the basic form [NP1, [NP2 PREDICATE]]. NP1 has topic-like function with respect to NP2 and both noun phrases have some claim to being subjects.
Errors in the translation of topic-comment structures

b. What error category accounts for the majority of errors? Does the percentage of errors differ in different groups of students from different years?
c. What pedagogical implications can be drawn for teachers to help students deal with such errors?

Subjects
Four groups of students total of 95 of the Department of English Languages and Literature of University of Social Sciences and Humanities of Ho Chi Minh City were chosen to be the informants of this study. Most of the students have studied English for nine years at high school. The dominant teaching method at the school is still the Grammar-Translation method and the translation aspect plays an important part. At the time when the test takes place, the students of Year 1 will not have attended the obligatory course in translation, the students from year 3 has taken 1 module (60 periods or 45 hours), the students of Year 3 have finished 3 modules and Year 4 have finished 5 modules in Vietnamese-English translation.

Data Collection and the Test
The subjects were asked to translate a 250-word text from Vietnamese into English in 180 minutes without being told which structure is being tested. The Vietnamese text is an extract from an article named ‘Mê Thảo-thơì vang bong’ of Tuổi Trẻ (The Youth), a well-known newspaper in Vietnam. The text is chosen because of its richness in sentence types: many different types of ‘authentic’ Vietnamese sentences can be found in the text, including minor sentences, i.e. elliptical sentences, sentences without either Topic or Comment, sentences in which the Topic is identical with the Subject, sentences in which the Topic is not identical with the Subject, sentences in which the Topic or Comment itself is another Topic-Comment structure. With such a variety in sentence types, the text promises to be a good environment for the author to discover what the most problematic structures are for the subjects when translating from Vietnamese to English. The text is also chosen because it contains traces of the topic-prominence of Vietnamese languages, which is hypothesized to pose a number of translation problems for students.

Content Analysis
Models of analysing
When the translators are concurrently the second language learners, the model of analysing errors and translation assessment must be based on the learning model, which is intended for teaching linguistics and training translation at the same time. Campell & Hale (2003: 218) mention the relationship between these factors in the following quotation:

Validity in interpreting and translation testing is tied up with knotty issues such as the nature of the competencies assessed, the models of learning underpinning educational programs, and the extent to which tests should reflect professional tasks. (Campell & Hale, 2003:218)

The present study suggests the following model of analysing errors, which is based on the learning model and context of translation training in Vietnam, where translation is
Pham Phu Quynh Na taught as a vocational skill for students at the universities, rather than as a tool for language teaching or an independent discipline by itself.

**ERROR CORPUS**

**COMPREHENSION ERRORS**

**LINGUISTIC ERRORS**

- Grammatical errors
- Syntactical errors
- Morphological errors
- Collocational errors
- Word Form

**TRANSLATION ERRORS**

- Omission
- Addition
- Inaccurate rendition of individual text items
- Distorted meaning of the source text
- Too free translation
- Too literal translation
- Pragmatic error
- Wrong lexical choice

**Figure 1: Models of analysing the error corpus**

**Data Analysis**

These errors were first detected and corrected by an American academic (Professor of History at National University of Singapore) who is highly competent in Vietnamese. He has used Vietnamese to teach Vietnamese history at Vietnamese National University and taught English as a second language at some Vietnamese universities. After the American examined the 95 translations to detect the errors, the author and two other Vietnamese ESL teachers double-checked the error correction of the native speaker to ensure that it is appropriate. The first step of correcting errors is to identify the linguistic errors, i.e. syntactic errors, grammatical errors, morphological errors. After all the linguistic errors had been sorted out, the 95 translations were re-examined from the communicative point of view to detect translation errors.

**Definition of Errors used in the Data Analysis:**

All errors will be divided into 3 types for this study: comprehension errors, linguistic errors and translation errors.
Errors in the translation of topic-comment structures

**Comprehension errors:** These errors happen when the learners misunderstand the syntax of a certain sentence of the source text, or misread a word and their translations were accordingly based on a misunderstood source text.

**Linguistic errors:** Under the umbrella term of ‘Linguistic errors’, there are following specific types of errors: grammatical errors, syntactic errors, morphological errors, collocational errors, incorrect word form.

  **Grammatical errors:** errors occur in the handling of word structure, including lack of agreement between subject and verb, incorrect verb tenses or verb forms, incorrect case of noun, pronouns, adjectives, and the use of an adjective when a verb is needed [American Translators Association (ATA) Framework for Standard Error Marking]. The word classes where grammatical errors happen include Noun, Pronoun, Verb, Conjunction, Adjective, Adverb, Preposition, Article and Determiner. The classification of error type was narrowed to the use of the following terms: Omissions, Additions, Inappropriate Combination, Inappropriate construction, Misordering or Inversion.

  **Syntactic errors:** errors made when learners have to handle any items larger than word, i.e. phrase, clause or sentence (James 1998). Errors in this category include sentence fragments, inappropriate use of relative clause, inappropriate sentence construction, improper modification, lack of parallelism, and misordering (unnatural word order) (ATA Framework for Standard Error Marking). The explanation of the nature of error type was narrowed to the use of the following terms: Omission, Addition, Inappropriate Combination, Inappropriate Choice, Inappropriate construction, Misuse, Misordering or Inversion, Misplacement

  **Morphological errors:** errors which involve a failure to comply with the norm in supplying any part of any instance of these word classes: six book*, aboli*shment… are noun morphology errors. (James 1998)

  **Collocational errors:** errors committed in the idiomatic usage of the target language. The errors in grammatical collocation (wrong use of preposition, etc.) are treated as grammatical errors, not as collocation error.

  **Incorrect word form** refers to cases where the word form is not correctly chosen, i.e. the learner uses the noun instead of the adjective to express in the target language. The root of the word is correct but the wrong form is used. This kind of error is put under the section of linguistic error, not translation error.

**Translation errors:** These errors happen when the students show the inability to render the original meaning of the source text into the target text. In other words, a sentence is considered to have ‘translation errors’ when it shows traces of distortions of the source text, although the sentence may be grammatically correct. In my analysis, translation errors include the omission of something essential to the meaning of the source text, the addition of unnecessary parts not intended in the original text, inaccurate renditions of some text items, the distortion at a certain level of meaning of the source text, too freely translated version or too literal translation’. Translation errors also include the pragmatic error and wrong lexical choice which will be defined as below.

---

Pham Phu Quynh Na

All of the errors mentioned in this study can be named with the general term ‘translational error’ in the sense that they happen during the process of translation. However, by classifying a certain error as a ‘translation error’, but not a ‘comprehension error’ or a ‘linguistic error’, the present author wants to refer to the error occurring in the process of transferring one sentence from the source text into the target text in the condition that the sentence grammar is acceptable. In other words, these ‘translation errors’ are more related the ‘transfer competence’ in the three-stage process of Nida (1964) and Nord (1992) as they are errors made in the process of transferring meaning. Accordingly they can be termed as ‘transfer errors’. However, to avoid the confusion between ‘language transfer errors’ in second language acquisition and the errors made in the ‘transfer process’ of translation, the term ‘translation errors’ is utilised instead of ‘transfer errors’.

Pragmatic error: errors made when learners produced a grammatically and semantically correct phrase/sentence but the use of the phrase/sentence is not appropriate with the communicative situation set out in the source text and target text.

Wrong lexical choice refers to the cases when the learner cannot find the right word to translate. To illustrate, the student translator might not have selected the most appropriate word among several that have similar (but not identical) meaning. This error is considered as a translation error because it may impair the comprehension of the readers.

Quantitative Analysis and Discussion

Table 1 reports the percentage of different types of errors in each year. The percentage of error reported in each cell will reflect the proportion of this type of error (i.e. syntactic error, grammatical error, etc.) compared to the overall number of errors of the same year. To illustrate, the percentage of Syntactic Errors in Year 2 column, which is 24.61%, means that syntactic errors occupies 24.61% of the total number of errors in Year 2. The comparison will be made between 1) distribution of different types of error in the same year (vertically) and 2) the distribution of the same type of error over the 4 years (horizontally)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Total</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic Errors</td>
<td>121</td>
<td>175</td>
<td>285</td>
<td>82</td>
<td>663</td>
<td>28.27%</td>
<td>24.61%</td>
<td>27.32%</td>
<td>19.85%</td>
</tr>
<tr>
<td>Grammatical Errors</td>
<td>124</td>
<td>230</td>
<td>335</td>
<td>157</td>
<td>836</td>
<td>28.97%</td>
<td>32.34%</td>
<td>31.16%</td>
<td>38.01%</td>
</tr>
<tr>
<td>Morphological Errors</td>
<td>44</td>
<td>75</td>
<td>127</td>
<td>52</td>
<td>298</td>
<td>10.28%</td>
<td>10.54%</td>
<td>17.86%</td>
<td>12.59%</td>
</tr>
<tr>
<td>Collocation Errors</td>
<td>29</td>
<td>37</td>
<td>66</td>
<td>23</td>
<td>155</td>
<td>6.77%</td>
<td>5.20%</td>
<td>6.30%</td>
<td>5.56%</td>
</tr>
<tr>
<td>Word Form</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>15</td>
<td>0.46%</td>
<td>0.56%</td>
<td>0.76%</td>
<td>2.42%</td>
</tr>
<tr>
<td>Comprehension Errors</td>
<td>19</td>
<td>32</td>
<td>49</td>
<td>22</td>
<td>122</td>
<td>4.34%</td>
<td>4.50%</td>
<td>4.69%</td>
<td>5.32%</td>
</tr>
<tr>
<td>Translation Errors</td>
<td>89</td>
<td>158</td>
<td>183</td>
<td>76</td>
<td>509</td>
<td>20.79%</td>
<td>22.22%</td>
<td>17.54%</td>
<td>18.40%</td>
</tr>
<tr>
<td>Total number of errors</td>
<td>428</td>
<td>711</td>
<td>1043</td>
<td>413</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Number of students</td>
<td>15</td>
<td>27</td>
<td>36</td>
<td>17</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

58
Errors in the translation of topic-comment structures

Distribution of different types of errors in the same year
The highest proportion of errors observed in all 4 years is grammatical errors (28.97% for Year 1, 32.34% for Year 2, 31.16% for Year 3, 38.01% for Year 4). The kind of errors that occupies the second highest percentage of errors in all 4 years is syntactic errors. This proves that linguistic competence is really one of the hindrances to the students. The errors that occupy the third highest proportion fall into the category of translation errors. In short, even though students may have problems with comprehension, they seem to have more difficulty in the second and third phase of translating process ‘transfer’ and ‘restructuring’ rather than ‘analysis’ phase.

Distribution of the same type of error over the 4 years
The table also helps to compare the distributions of one single type of error across four groups of students. Regarding grammatical errors, Year 4 students had the highest percentage of errors (38.01%) compared to all other 3 years (28.97%, 31.16%, 32.34%). In the field of morphological errors, Year 3 students had the highest percentage distribution (17.83%). Year 2 students had the highest percentage of translation errors (22.22%) among 4 years and Year 1 students had the highest number of errors in collocation handling (6.77%) and syntactic errors (28.27%).

The fact that grammatical errors occupy the largest percentage in the total number errors of Year 4 does not necessarily mean that Year 4 students made more grammatical errors than students of other years. It may simply mean that Year 4 students seemed to focus more on the retaining the context’s meaning, rather than on the processing of each individual sentence. Knowing that word-by-word translation is not the correct way of translating, they tended to concentrate more on the ‘transferring’ phase rather then the ‘restructuring’ phase. Therefore, they seemed to ignore the specific task of guaranteeing the correct grammar of each individual sentence. Year 4 students also had high percentage of comprehension errors compared to other years, probably because they had so much self-confidence in their comprehension ability of Vietnamese and therefore took no notice of the details of the test.

It is understandable that Year 1 students made more syntactic errors and collocation errors compared to three other years. Year 1 students seemed not to have high linguistic competence, which hindered them from transferring the complex syntactic structure of Vietnamese sentences to fit the subject-predicate structures of English sentences. Also Year 1 students may have limited knowledge of collocation and they tend to make errors in this field. Year 1 students may have known that their linguistic competence is limited and were likely to be more careful in reading and comprehending the text as well as in the translating of each individual sentence.

Year 2 students made more errors in the process of transferring sentence from Vietnamese into English, not in the process of comprehending/analysing the text or restructuring the sentence. Possibly because Year 2 is the first year when the students have just been taught Translation at the university, they want to attempt more strategies in their translation. However, their still limited knowledge prevented them from having exact translational strategy.

---

3 Nida’s model of translation
<table>
<thead>
<tr>
<th>SENT.NO</th>
<th>ERROR TYPE</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>Frequency</td>
<td>6</td>
<td>148</td>
<td>70</td>
<td>26</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>2.0%</td>
<td>17.7%</td>
<td>10.6%</td>
<td>16.8%</td>
<td>33.3%</td>
<td>2.5%</td>
</tr>
<tr>
<td>2.00</td>
<td>Frequency</td>
<td>6</td>
<td>20</td>
<td>28</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>2.0%</td>
<td>2.4%</td>
<td>4.3%</td>
<td>1.9%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3.00</td>
<td>Frequency</td>
<td>39</td>
<td>52</td>
<td>34</td>
<td>1</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>13.1%</td>
<td>6.2%</td>
<td>5.2%</td>
<td>0.6%</td>
<td>0.0%</td>
<td>6.6%</td>
</tr>
<tr>
<td>4.00</td>
<td>Frequency</td>
<td>66</td>
<td>42</td>
<td>52</td>
<td>3</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>22.1%</td>
<td>5.0%</td>
<td>7.9%</td>
<td>1.9%</td>
<td>6.7%</td>
<td>16.5%</td>
</tr>
<tr>
<td>5.00</td>
<td>Frequency</td>
<td>5</td>
<td>20</td>
<td>68</td>
<td>0</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>1.7%</td>
<td>2.4%</td>
<td>10.3%</td>
<td>0.0%</td>
<td>6.7%</td>
<td>8.8%</td>
</tr>
<tr>
<td>6.00</td>
<td>Frequency</td>
<td>14</td>
<td>32</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>4.7%</td>
<td>3.8%</td>
<td>1.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>3.3%</td>
</tr>
<tr>
<td>7.00</td>
<td>Frequency</td>
<td>5</td>
<td>21</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>1.7%</td>
<td>2.5%</td>
<td>0.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>8.00</td>
<td>Frequency</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>7.0%</td>
<td>1.8%</td>
<td>3.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>9.00</td>
<td>Frequency</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>7.0%</td>
<td>2.5%</td>
<td>3.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>10.00</td>
<td>Frequency</td>
<td>3</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>1.0%</td>
<td>2.0%</td>
<td>3.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>11.00</td>
<td>Frequency</td>
<td>8</td>
<td>97</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>2.7%</td>
<td>11.6%</td>
<td>1.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>12.00</td>
<td>Frequency</td>
<td>25</td>
<td>65</td>
<td>61</td>
<td>85</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>8.4%</td>
<td>7.8%</td>
<td>9.3%</td>
<td>54.8%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>13.00</td>
<td>Frequency</td>
<td>2</td>
<td>60</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>7.0%</td>
<td>7.2%</td>
<td>2.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>14.00</td>
<td>Frequency</td>
<td>16</td>
<td>81</td>
<td>39</td>
<td>34</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>5.4%</td>
<td>9.7%</td>
<td>5.9%</td>
<td>21.0%</td>
<td>20.0%</td>
<td>1.7%</td>
</tr>
<tr>
<td>15.00</td>
<td>Frequency</td>
<td>25</td>
<td>39</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>8.4%</td>
<td>5.7%</td>
<td>8.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>16.00</td>
<td>Frequency</td>
<td>1</td>
<td>6</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>0.03%</td>
<td>7.0%</td>
<td>4.7%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>17.00</td>
<td>Frequency</td>
<td>69</td>
<td>49</td>
<td>84</td>
<td>3</td>
<td>1</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>23.2%</td>
<td>5.9%</td>
<td>12.8%</td>
<td>1.9%</td>
<td>6.7%</td>
<td>64.5%</td>
</tr>
<tr>
<td>18.00</td>
<td>Frequency</td>
<td>4</td>
<td>51</td>
<td>94</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>1.3%</td>
<td>6.1%</td>
<td>14.3%</td>
<td>0.0%</td>
<td>26.7%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

| Total   | Frequency  | 298 | 836 | 658 | 155 | 15 | 121 | 526 | 2609 |
|         | Percentage | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|

1.1 Morphological Errors 1.2 Grammar Errors 1.3 Syntactic Errors 1.4 Collocation Errors 1.5 Word Form 2 Comprehension Errors 3 Translation Errors

Table 2 shows the distribution of different types of error (respectively, morphological errors, grammar errors, syntactic errors, collocation errors, word form, comprehension errors, translation errors) across the 18 sentences. That is to say, the
Errors in the translation of topic-comment structures

Rows represent the sentences and the columns represent the types of error; each number appearing in the box will show the corresponding percentage of one specific type of error in that sentence. The percentage in each row will be calculated by the frequency of error in each sentence divided by the overall number of errors of the same type, considered as 100% at the end of each column. The numbers highlighted in red show the errors that account for the highest percentage. Among all morphological errors of 18 sentences, the morphological errors have the highest frequency and proportion in sentence 17 (23.2%) and sentence 4 (22.1%). The highest percentage of grammatical errors across the 18 sentences are seen in sentence 1 (17.7%) and sentence 11 (11.6%). Among the 8 sentences, syntactic errors appear with highest frequency and percentage distribution in sentences 18 (14.3%) and 17 (12.8%). Collocation errors are seen with the largest percentage in sentences 12 (54.8%) and 14 (21.9%). Among all morphological errors of 18 sentences, the morphological errors have the highest frequency and proportion in sentence 17 (23.2%) and sentence 4 (22.1%). The highest percentage of grammatical errors across the 18 sentences are seen in sentence 1 (17.7%) and sentence 11 (11.6%). Among the 8 sentences, syntactic errors appear with highest frequency and percentage distribution in sentences 18 (14.3%) and 17 (12.8%). Collocation errors are seen with the largest percentage in sentences 12 (54.8%) and 14 (21.9%). Students have the greatest problems with the formation of word in sentence 18 (26.7%) and sentence 1 (33.3%). The largest distribution of comprehension errors is seen in sentences 17 (64.5%) and 4 (16.4%). Translation errors have a rather even dispersion in the 18 sentences. This shows that although students may have certain text-specific or linguistic-specific problems posed by some syntactic structure, collocation or comprehension points in some sentences, translation skills are still problematic for them. Translation errors are the most dominant in two sentences, sentence 18 (11.7%) and sentence 6 (8.6%).

Findings

1. Among all types of errors, grammatical errors represent the largest proportion. Next come errors of syntax, translation, word form, comprehension and collocation. Although the comprehension errors account for only a small proportion, they seem to have serious consequences as they may lead to mistranslation and thus misunderstanding to the readers.

2. Among all types of syntactic errors, the four main types that account the highest percentage at the level of sentence structure are Inappropriate Sentence Construction (19.14%), Omission of Main Verb (13.2%), Misuse of Relative Clause (7.01%), and Not Parallel Combination (7.01%). These four types of syntactic errors can be said to be more or less attributable to the Topic-Comment Structure of Vietnamese.

3. In sentences whose topic-comment structure is remarkably different from the subject-predicate structure, students did have problems in locating the subject. However, they seem to have more problems in the task of handling the relationship between the subject and the verb: they are puzzled to recognise the difference between verbal predicates (in the form of a verb) and substantival predicates (in the form of an adjective) in Vietnamese and they often omit the verbs in the target text versions. They also have difficulty in handling the relative clause in sentences having two topic-comment structures and in building the sentence in an appropriate way.

4. The syntactic errors are observed in the specific task of translating subjects, to see whether students can handle the subject in the context where the initial constituent at the beginning of each sentence is always the topic, which may or may not coincide with the subject. Five types of syntactic errors occur in this task of locating and rendering the subject into the target text: Omission of Subject, Repetition of Subject, Inappropriate Choice of Subject, No Logical Connection Between Subject and Predicate, Inappropriate Connection Between Subject and Passive Verb. Observation shows that Omission of Subject is the most frequent type of error, followed by
Repetition of Subject and No Logical Connection Between Subject and Predicate. However, it is worth noting that these errors are not the most frequent among the various kinds of syntactic errors in this study.

5. Among four years, Year 1 make more syntactic and collocation errors compared to three other years. Year 4 make more grammatical errors than other years. This fact does not necessarily mean that the linguistic competence of Year 4 is lower than other years. It can be explained by the fact that Year 4 seem to focus more on the context’s meaning, rather than on the processing of each individual item in the sentence. Year 3 have the highest percentage distribution of morphological errors. Year 2 make more errors in the process of transferring the sentence from Vietnamese into English, not in the process of comprehending the text.

Pedagogical Implication
The impression extracted from the data is that the students have a rather limited knowledge of both Vietnamese and English. The lack of a thorough knowledge of Vietnamese grammar imposed a certain level of difficulty to students. Their failure to master the basic characteristics of Vietnamese as a topic-prominent language and to understand the structural difference between Vietnamese and English impeded students from choosing a suitable equivalent for a certain Vietnamese syntactic structure. The students also seemed to take up lexical and grammatical knowledge of English expressions but not their pragmatic and discoursal use. As a consequence, the most regular mistranslations by the Vietnamese students were the use of unconnected combination of some words or phrases. This tended to produce sentence forms that were very cumbersome and unnatural in English. The students were not well-equipped with the knowledge of the function of the English language. Such knowledge would have helped them to maintain the focus (or prominence) of information presented when translating from the one language (Vietnamese) to the other (English). In some instances where the students were proficient enough to select the correct equivalent structure in English, they still had problems locating the subject that best conveyed the meaning in the source sentence without changing its focus of information.

To deal with the phenomena presented above, the following strategies are recommended in the teaching of Vietnamese-English translation:

1. In the program of Vietnamese-English translation, a certain number of hours should be spared for teaching Vietnamese grammar and the special role of topic-prominent structures in Vietnamese language. Recognising these typological differences in syntactic structures between Vietnamese and English, students can become more alert when choosing the suitable structure in translation.

2. In teaching any new word or phrase, the teacher should provide the students not only its semantic use but also its pragmatic and discoursal use.

3. It is a tradition in Vietnamese universities that, in translation classes, 200-word texts in the source language are usually given to students and the teacher will work with the students on that text. Although this makes it easy for students in terms of their choice of vocabulary, it fails to provide students with a good resource of available ‘equivalent structures’ they can use when faced with a specific type of Vietnamese topic-prominent language. It is recommended that, instead of being given 200-word text, students should be given 10 sentences of the same syntactic structures in Vietnamese on different topics in order that they can master the way of handling
Errors in the translation of topic-comment structures

subjects in a given structure. It may be argued that this way of teaching may encourage word-for-word translation. However, for learners of English with very limited competency, this way of teaching at least can help them to handle the text correctly, before they are competent enough to make it perfect and natural in the whole text.

4. This study suggests the use a choice network analysis based on the error corpus of students. According to Campbell & Hale (2003), a choice network is in effect a psycholinguistic model of the options facing the translator, based on the evidence of the target texts of groups of subjects translating the same text. The suggestion is that different translations from many students from the same source text should be collected. These different ways of translating will be listed and systematised into a ‘choice network’, which can be used and referred to any time the same source text is used again for teaching purpose. This way, students of different classes or from different years can avoid the possible errors and discuss all available ways of translating a certain structure before they can choose the most suitable to convey the source text.

Conclusion

The study investigated three questions: a) whether the students have problem translating sentences in which the Topic of the topic-comment structure does not coincide with the Subject, b) what kinds of errors students made in translating this specific type of structure and c) whether the percentage of errors differ in different groups of students from different years. The number of errors found in the data may not be large enough to substantiate the fact that Vietnamese EFL students always have a great deal of difficulty in handling the typological differences between Vietnamese and English. However, there is enough evidence to suggest that these errors still need to be identified and collected during translation training. Grammatical errors still occupy the highest percentage errors among all other types. The study has identified some typical syntactic errors in the Vietnamese-English translation caused by the influence of the Vietnamese topic-comment structures. The most common errors in the translation of these structures include Omission of Subject, Repetition of Subject, Inappropriate Choice of Subject, No Logical Connection Between Subject and Predicate, Inappropriate Connection Between Subject and Passive Verbs. Some strategies were suggested that can be applied to prevent the possible problems arising from translating topic-prominent structures of Vietnamese into English. Although the percentages of errors differ in different groups of students, there is no clear evidence to show that this variation is the direct result of the discrepancy in their language competence.

References


Pham Phu Quynh Na


Errors in the translation of topic-comment structures


Appendix A

ELICITATION TEST

Translate the following passages into English:

(1) Laø keû töøng chaêm chuù doõi theo cuoäc ‘haønh trình thai saûn’ cuûa boä phim, toâi meâ maån vói Mé Thaûo (giaù chæ ñaët teân phim laø Mé Thaûo thoâi thì nghe giaûn dò hôn vaø ‘ñaõ’ hôn).

(2) Tröôùùc heát laø meâ maån vôùi nhöõng hình aûnh ‘raët ngoân ngöõ ñieän aûnh’.


Title: Tœ Çhaø Dœn dœn Mé Thaûo - ñøi vang bœng

Author: Nguyen Duy

Source: Tuoåi Treû Chuû Nhaät (Sunday Youth)
Number 38-2002 Date 29-9-2002
Appendix B

STRUCTURAL CONFIGURATION OF 18 SENTENCES IN THE ELICITATION TASK

(1) (Tôi) Là kẻ từng chăm chú đổi theo cuộc ‘hành trình thai sản’ của bộ phim, me mám với Me Thảo.

(2) Trước hết là (Tôi) me mám với những hình ảnh ‘rất ngôn ngữ’ diễn ảnh.

Cảnh đốt bàn ghế, tiếng nổ của đồ đạc và phát súng bắn vào con búp bê khiến người xem lạnh giá.

Scene burn furniture sound explode of furniture and the shoot shoot reposition the doll make audience cold

làm ta sồn gai ốc.
make pronoun creep goose-flesh

Cảnh thả đen trời (người ta/người xem..) tưởng như chỉ có ma thuật mỗi làm nỗi

Scene drop light sky (people/audience) seem only magic in order to do

Cảnh quay nào (người ta/người xem) cũng thấy mồ hôi.
Scene any (people/audience) also see sweat
Errors in the translation of topic-comment structures

(7) Mồ hôi của ý nghĩ.
Sweat of thought

TOPIC (NO COMMENT)

(12) Trong trí nhớ tôi chưa có một phim nào của nước Việt ta.
In memory I not yet have one film any of Vietnamese

RANGE TOPIC TOPIC (EXISTENTIAL CLAUSE)

EXISTENTIAL CLAUSE (THERE..)
dựng thực hiện công phu và đạt chuẩn như vậy
be/get carry out elaborate and obtain standard like that.

(13) (Tôi)
Lại mê mẩn với những nhân vật có tên và không tên.
(1) again be charmed with plural marker character have name and no name.

NULL TOPIC COMMENT

NULL SUBJECTPREDICATE

(14) Trong nhiều nhân vật phụ thành công, ông bố già rất sinh động và cảm động
In many character secondary successful man servant old very lively and touching

RANGE TOPIC TOPIC COMMENT 1

SUBJECT PREDICATE 1

(ông) xứng đáng được bầu là vai phụ xuất sắc nhất
(he) deserving be/get nominate be character secondary excellent best

NULL TOPIC 2 COMMENT 2

NULL SUBJECTPREDICATE 2

(15) Nhân vật chính nào cũng mỗi người một vẻ, độc đáo, đầy thân phận,
Character main any also each person one style original full condition

TOPIC COMMENT

TOPIC 1 COMMENT 1

SUBJECT PREDICATE 1
Phạm Phú Quỳnh Na

(16) Nguyễn hào hoa, hào hiệp, lập dị ra mắt, (anh) rõ là anh diện chịu bất đặc chí, khắt khửng.

(17) Tam, cây đàn nguyệt tuyệt vời tới mức tuyệt vọng, u uẩn, bề tắc mà nghĩa khí,
cử chỉ thật chính xác với tâm trạng.

(18) Tôi, cô dâu hát hồng nhan bạc phận, phiêu dat mà cao sang, da ñình mà chung tình,
Errors in the translation of topic-comment structures

vừa quyền rủ vừa nghiêm trang, nhỏ nhoi mà không hèn mọn
both attractive and serious petite but not lowly

khó có ai nhắm vai (Tô) hay hôn Thùy Nga
difficult have anyone play role (Tô) good more Thùy Nga

TOPIC (EXISTENTIAL CLAUSE) COMMENT
SUBJECT-RAISING CONSTRUCTION
Nominalization and Categorization of Verbs in Thai

Amara Prasithrathsint
Chulalongkorn University
Amara.Pr@Chula.ac.th

1. Introduction
Nominalization refers to “the process of forming a noun from some other word class (e.g. red-ness) or (in classical Transformational Grammar especially) the derivation of a noun phrase from an underlying clause (e.g. His answering of the letter... from He answered the letter)” (Cystal 1991: 232-234).

This study deals with the first type of nominalization, which is labeled as “lexical nominalization” by Givón (1984: 498). It is a process of lexical formation or word formation. In Thai, lexical nominalization means the process of deriving nouns from verbs, which can be done in several ways. The present study focuses only on deriving abstract nouns from verbs. This process may be specially labeled as “abstract nominalization” and the derived nouns as “abstract nominals” (see Prasithrathsint 1997).

Abstract nominals in Thai are marked by either kaan- or khwaam-; for example:

1. kin ‘to eat’ → kaan-kin ‘eating’
2. phāṭhanaa ‘to develop’ → kaan-phāṭhanaa ‘development’
3. sadeey ‘to perform’ → kaan-sadeey ‘performance’
4. suk ‘to be happy’ → khwaam-suk ‘happiness’
5. yūṭitham ‘to be just’ → khwaam-yūṭitham ‘justice’
6. ciṭçay ‘to be sincere’ → khwaam-ciṭçay ‘sincerity’

According to Prasithrathsint (1997), the nominalizers kaan- and khwaam-developed from the lexical nouns kaan ‘work, affair, matter’, and khwaam ‘sense or substance of a matter; legal case, lawsuit’, respectively. The study also shows that kaan-nominalization emerged after khwaam-nominalization. In the Sukhothai period (the 13th century), three incidences of khwaam-nominalization were found. The kaan-nominalization, which is the most general pattern today, emerged in the late Ayutthaya period, i.e. around four centuries later, but became productive in the 20th century.

There are interesting questions concerning these two nominalizers in Thai; for example: "Why does Thai need two nominalizers?" "How are they different?" Are they complementarily distributed among the verbs in Thai? Prasithrathsint (1996) shows that...

---

1 I am grateful to Natchanan Yaowapat who presented the earlier version of this paper at SEALS XV on my behalf. Thanks also go to those who gave comments, especially Pittayawat Pittayaporn.

2 The other type of nominalization is “grammatical nominalization”, which refers to “the process via which a prototypical verbal clause—either a complete sentence (including the subject) or a verb phrase (excluding the subject) is converted into a noun phrase.” (Givón 1984: 498)
the two nominalizers differ in their occurrence in speech styles; i.e., kaan- nominals tend to occur more frequently in academic writing and khwaam- nominals in description and narration.

It is important to note that choosing kaan- or khwaam- to add to a verb is not at all arbitrary. There seem to be some rules that control the choice. However, so far there has been no attempt to explain precisely when to use kaan- and when to use khwaam-. On the contrary, there are only observations; for instance, Haas (1964: 29, 82) notes that when placed before active verbs to form noun derivatives, kaan- can almost always be rendered in English by the suffix -ing and sometimes by -ion, -sion, -al, -ance/-ence, and -ment. As for khwaam-, she says that it is usually rendered in English by -ness, -ity, -ment, -ance/ence, -(t)ion, -(t)ude, -ure, -y, -ery, -ship, -dom, and sometimes -ing. Yaemmadda (1987) suggests that khwaam- occurs with verbs, adjectives, and nouns, but kaan- with verbs only. He also adds that generally nominals with kaan- are similar to gerunds in English and that kaan- signifies movement whereas khwaam- signifies stability.

Suggesting that they are differentiated by types of verb, Morev (2004) states that kaan- is used for converting verbs, particularly action verbs, into abstract nouns, and that khwaam- is used for turning non-action verbs; i.e., verbs denoting states, feelings, qualities and properties into abstract nouns. He also suggests that in some cases both can be used with the same verb, but the meaning of the derived nominals will be different, e.g. kaan-taay ‘dying’, ‘death’ (a name of process) vs. khwaam-taay ‘death’.

As can be seen from the above observations, kaan- and khwaam- overlap to a considerable extent. Nevertheless, it can be generalized that kaan- is used with action verbs and khwaam- with non-action or stative verbs. However, this generalization is not completely valid because there are counterexamples. In actuality, some non-action verbs, which should be nominalized with khwaam-, are found to be nominalized with kaan-; for example:

7. ráprúu ‘to perceive’ \(\rightarrow\) kaan-ráprúu ‘perception’  
     * khwaam-ráprúu

8. dâyyin ‘to hear’ \(\rightarrow\) kaan-dâyyin ‘hearing’  
     * khwaam-dâyyin

9. dâyklin ‘to smell’ \(\rightarrow\) kaan-dâyklin ‘smelling’  
     * khwaam-dâyklin

10. rianruú ‘to learn’ \(\rightarrow\) kaan-rianruú ‘learning’  
     * khwaam-rianruú

11. trâtsaruú ‘to be enlightened’ \(\rightarrow\) kaan-trâtsaruú ‘enlightenment’  
     * khwaam-trâtsaruú

12. damrogyûú ‘to exist’ \(\rightarrow\) kaan-damrogyûú ‘existence’  
     * khwaam-damrogyûú

13. wikhr’ ‘to analyze’ \(\rightarrow\) kaan-wikhr’ ‘analysis’  
     * khwaam-wikhr’

14. phakph’ú ‘to rest’ \(\rightarrow\) kaan-phakph’ú ‘rest’  
     * khwaam-phakph’ú

---

3 In this study, adjectives in Thai are considered to be a sub-class of verbs called “adjectival verbs”, because syntactically they behave like verbs.
Nominalization and categorization of verbs in Thai

In addition, some verbs are found to occur with both kaan- and khwaam-; for example:

15. riłprɔ́m ‘to hurry’ → kaan-riłprɔ́m ‘hurry’
    → khwaam-riłprɔ́m ‘hurry’
16. klomkliin ‘to assimilate’ → kaan-klomkliin ‘assimilation’
    → khwaam-klomkliin ‘assimilation’
17. riǔąmmii ‘to cooperate’ → kaan-riǔąmmii ‘cooperation’
    → khwaam-riǔąmmii ‘cooperation’
18. klińkwày ‘to move’ → kaan-klińkwày ‘movement’
    → khwaam-klińkwày ‘movement’
19. chuàyìka ‘to assist’ → kaan-chuàyìka ‘assistance’
    → khwaam-chuàyìka ‘assistance’
20. anukhrɔ́ ‘to subsidize’ → kaan-anukhrɔ́ ‘subsidy, subsidization’
    → khwaam-anukhrɔ́ ‘subsidy, subsidization’
21. taay ‘to die, to be dead’ → kaan-taay ‘dying, death’
    → khwaam-taay ‘dying, death’
22. BUT kɔ̀ɔ ‘to be born’ → kaan-kɔ̀ɔ ‘birth’;
    → *khwaam-kɔ̀ɔ ‘birth’

The examples above show irregularities in deriving abstract nouns from verbs in Thai. However, those irregularities may be superficial, and there may be some explanations for their occurrences. This study aims to provide such explanations. The above-mentioned suggestions given by Morev (2004) serve as hypotheses on the difference between the two nominalizers with reference to verbal categories in Thai.

Therefore, this study attempts to find out which types of verbs are nominalized by kaan- and which by khwaam- and in what ways they are differentiated.

The data on which the analysis is based are 5,600 verbs in Thai, which consist of kaan-/khwaam- nominalized verbs found in a two-million-word corpus of current Thai, and verbs taken from the Royal Institute Dictionary of Standard Thai (1999).

2. The frequencies of three patterns of nominalization in Thai

Based on an analysis of the frequency of the occurrences of verbs with kaan-, khwaam-, or both kaan- and khwaam- in the data, it can be summarized that abstract nominalization in Thai can be divided into three patterns, and that the kaan- pattern is the most frequent. Next is the khwaam- pattern, and the kaan-/khwaam- pattern is the least frequent (see Table 1). The result confirms the findings in Prasithrathsint (1997) that kaan- is the most general nominalizer in present-day Thai.

<table>
<thead>
<tr>
<th>Patterns of nominalization</th>
<th>Frequency (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kaan-</td>
<td>51</td>
</tr>
<tr>
<td>khwaam-</td>
<td>33</td>
</tr>
<tr>
<td>kaan-/khwaam-</td>
<td>16</td>
</tr>
<tr>
<td>total</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: The frequencies of the three patterns of nominalization in Thai
3. Three categories of verbs in Thai
An analysis of each of the nominalized verbs in the data and its syntactic-semantic environments enabled me to divide the verbs into three categories according to their occurrences in each of the nominalization patterns mentioned in Section 2. The three categories of verbs are labeled here as perceptible, imperceptible, and balanced verbs.

3.1 Perceptible verbs
Perceptible verbs occur with kaan- only. They represent concepts that are easy to understand or characterize. They may be action or non-action verbs, and their meanings are perceivable due to certain common features or standards shared by native speakers.

**Examples:**

23. kõm ‘to bend’ → kaan-kõm; *khwaam-kõm
24. klaaw ‘to state’ → kaan-klaaw; *khwaam-klaaw
25. kɔt ‘to be born’ → kaan-kɔt; *khwaam-kɔt
26. kin ‘to eat’ → kaan-kin; *khwaam-kin
27. khäft ‘to sell’ → kaan-khaft; *khwaam-khaft
28. khuy ‘to talk’ → kaan-khuy; *khwaam-khuy
29. khamooy ‘to steal’ → kaan-khamooy; *khwaam-khamooy
30. khçípkhun ‘to thank’ → kaan-khçípkhun; *khwaam-khçípkhun
31. cãtkaan ‘to manage’ → kaan-cãtkaan; *khwaam-cãtkaan
32. khàwruàm ‘to participate’ → kaan-khàwruàm; *khwaam-khàwruàm
33. khońphòp ‘to discover’ → kaan-khońphòp; *khwaam-khońphòp
34. nèñam ‘to introduce’ → kaan-nèñam; *khwaam-nèñam
35. pàırüup ‘to reform’ → kaan-pàięrüup; *khwaam-pàięrüup
36. àihibaay ‘to explain’ → kaan-àihibaay; *khwaam-àihibaay
37. àphípraay ‘to discuss’ → kaan-àphípraay *khwaam-àphípraay
38. hùar ‘to laugh’ → kaan-hùar *khwaam-hùar
39. hày ‘to give’ → kaan-hày *khwaam-hày
40. yių ‘to stop’ → kaan-yių; *khwaam-yių
41. hàaycay ‘to breathe’ → kaan-hàaycay; *khwaam-hàaycay
42. sànya ‘to promise’ → kaan-sànya; *khwaam-sànya
43. sàns ‘to praise’ → kaan-sàns; *khwaam-sàns
45. sójps ‘to promote’ → kaan-sójps; *khwaam-sójps
46. ánúyáát ‘to permit’ → kaan-ánúyáát; *khwaam-ánúyáát

As can be seen from examples 23-46 and also examples 7-14 in section 1, the concepts represented by the verbs are discernible no matter whether they are associated with physical or mental activities. In terms of syntactic characteristics, they can be transitive or intransitive. In terms of semantic features, they can be verbs of activity, process, bodily sensation, transitional event, etc.

3.2 Imperceptible verbs
Imperceptible verbs occur with khwaam- only. In contrast with perceptible verbs, imperceptible verbs represent concepts that are difficult to understand or characterize.
Nominalization and categorization of verbs in Thai

Their meanings may vary among native speakers. In accordance with Morev’s observations, these verbs denote values, feelings, qualities and properties.

Examples:

47. klom 'to be round' → khwaam-klom; *kaan-klom
48. klāā 'to be brave' → khwaam-klāā; *kaan-klāā
49. kwāā 'to be wide' → khwaam-kwāā; *kaan-kwāā
50. suāā 'to be high, tall' → khwaam-suāā; *kaan-suāā
51. sadiāak 'to be convenient' → khwaam-sadiāak; *kaan-sadiāak
52. khayaān 'to be diligent' → khwaam-khayaān; *kaan-khayaān
53. khoā 'to be bitter' → khwaam-khoā; *kaan-khoā
54. campen 'to be necessary' → khwaam-campen; *kaan-campen
55. cufluci 'to be fussy' → khwaam-cufluci; *kaan-cufluci
56. chala ›at 'to be intelligent' → khwaam-chala ›at; *kaan-chala ›at
57. No 'to be stupid' → khwaam-No; *kaan-No
58. rew 'to be quick' → khwaam-rew; *kaan-rew
59. riap 'to be smooth' → khwaam-riap; *kaan-riap
60. riapţay 'to be simple' → khwaam-riapţay; *kaan-riapţay
61. mēettaa 'to be kind' → khwaam-mēettaa; *kaan-mēettaa
62. māncay 'to be confident' → khwaam-māncay; *kaan-māncay
63. ruţriskaj 'to prosper' → khwaam-ruţriskaj; *kaan-ruţriskaj
64. caydam 'to be merciless' → khwaam-caydam; *kaan-caydam
65. caysò 'to be easily scared' → khwaam-caysò; *kaan-caysò
66. caytxŏn 'to be soft-hearted' → khwaam-caytxŏn; *kaan-caytxŏn
67. cayhāay 'to be shocked' → khwaam-cayhāay; *kaan-cayhāay
68. khřim 'to be solemn' → khwaam-khřim; *kaan-khřim

Examples 47-68 show that imperceptible verbs in Thai tend to be adjectival verbs, which are rendered in English by adjectives. Syntactically, they are intransitive verbs. Semantically, they are verbs of quality. Their meanings are not easy to discern because their semantic components seem to vary among speakers. However, the degree of quality represented by khwaam- nominals derived from an imperceptible verb can be measured by certain values or standards and perceived as a volume or amount. Thus, when one says khōw mii khwaam-campen (he has necessity) or khōw mii khwaam- mēettaa (he has kindness), it means something like ‘he has/possesses an amount of necessity’ and ‘he has/possesses an amount of kindness.’ This is evaluated or judged by the speaker’s own values.

3.3 Balanced verbs

Verbs of the third category are the most problematic; they are ambivalent as to their ability to occur with either kaan- or khwaam-. They are labeled here as “balanced verbs”, which seem to have characteristics of both perceptible and imperceptible verbs. Actually, the fact that this category exists seems to strengthen the belief that kaan- and khwaam- have certain inherent meanings. Examples of balanced verbs are 15-22 above and the following examples (69-90).
Amara Prasithrathsint

69. kratirilón ‘to be enthusiastic’
70. klokmkitín ‘to assimilate’
71. klua ‘to be afraid’
72. kròot ‘to be angry’
73. khāwcay ‘to understand’
74. khoikoon ‘to cheat’
75. khātwāj ‘to expect’
76. rak ‘to love’
77. mëap ‘to be silent’
78. caretēxpto ‘to grow’
79. chămān ‘to be sure’
80. taay ‘to die, to be dead’
81. thãmtua ‘to be modest’
82. thaarun ‘to be cruel’
83. naphī ‘to respect’
84. pramāat ‘to be careless’
85. plīnpleèng ‘to change’
86. thāathaay ‘to challenge’
87. maw ‘to be drunk’
88. sōncay ‘to be interested’
89. phāphē ‘to be generous’
90. òöüät ‘to boast’

The following sentences show some balanced verbs in context. It will be interesting to see that the nominals with kaan- and those with khwaam- have different meanings.

91. taay ‘to die, to be dead’

- kaan-taay yāay-saçoop pen thū-präatthanāa khājñ thēk khon
dying like-peace be place-desire of every person
‘Peaceful dying is desirable to everyone.’ (Here kaan-taay means the act of
dying.)

- chān dāay khāw kaan-taay khājñ khāw lēēw
I get news death of he already
‘I have heard the news of his death.’ (kaan-taay refers to an objective incident in
life.)

- khwaam-taay pen siŋ nāa-klua
death be thing fearful
‘Death is a fearful thing.’ (The speaker means ‘the body of death’)

92. klua ‘to be afraid’

- kaan-klua phū pen siŋ thū athībaay dāay yāak
fear ghost be thing that explain can difficult
Nominalization and categorization of verbs in Thai

‘Fear of ghosts is something difficult to explain.’ (generally speaking)

- **thuk**  **khon**  **mii**  **khwaam-klua**
  every person have fear before taking an exam
  ‘Everyone has fear (an amount of fear) before taking an exam.’

93. **krôot**  ‘to be angry’

- **kaan-krôot**  **at**  **tham-hây**  **khun**  **kêe**  **rew**
  Getting angry may cause you to be old fast
  ‘Being angry may cause you to get old fast.’ (generally speaking)

- **chân**  **kamlaŋ**  **phayaayaam**  **râŋâp**  **khwaam-krôot**
  I PRO try stop anger
  ‘I am trying to stop my anger.’ (specific meaning)

94. **khâwcay**  ‘to understand’

- **kaan-khâwcay**  **phuu-îın**  **pen**  **siŋ**  **campen**
  understanding others be thing necessary
  ‘Understanding others is a necessity.’ (general meaning)

- **khârpîkhun**  **sâmrâp**  **khwaam-khâwcay**
  thank you for understanding
  ‘Thank you for understanding.’ (specific meaning)

95. **khoîkooŋ**  ‘to cheat’

- **kaan-khoîkooŋ**  **tham-hây**  **khâw**  **tit-khuk**
  cheating cause him imprisoned
  ‘Cheating caused him to be imprisoned.’ (an act of cheating)

- **khwaam-khoîkooŋ**  **pen**  **laksanâ**  **dêñ**  **khâpîkhâw**
  cheating be characteristic prominent of he
  ‘Cheating is a prominent characteristic of his.’ (the quality of cheating)

96. **khâutwaŋ**  ‘to expect’

- **kaan-khâutwaŋ**  **at**  **tham-hây**  **khun**  **phîtwaŋ**
  expectation may cause you be disappointed
  ‘Expectation may cause you to be disappointed.’ (general meaning)

- **chân**  **mii**  **khwaam-khâutwaŋ**  **sûŋ**  **kîaw-kâp**  **yœaŋ**  **khâpîkhâw**
  I have expectation high about work of he
  ‘I have high expectation about his work.’ (specific meaning)
97.  rak  ‘to love’

- kaan-rak  khon-im  tham-hāy  raw  mii  khwaam-sūk
  loving  others  cause  we  have  happiness
  ‘Loving others makes us happy.’ (general meaning)

- dēk  thū  kāat  khwaam-rak  mak  mii  panhāa
  child  that  lack  love  tend  to  have  problem
  ‘A child who needs love tends to have problems.’ (amount of love)

4. Conclusion
In the preceding pages I have attempted to show how kaan- and khwaam- nominalizers are different in their inherent meanings and their distribution with verbs. It is found that verbs in Thai can be divided into three categories according to their ability of occurring with kaan- or khwaam-; i.e., perceptible verbs, imperceptible verbs and balanced verbs.

Perceptible verbs are the majority of verbs in Thai. They are either transitive or intransitive, action or non-action, physical or mental. They are different from imperceptible verbs in that they represent concepts that are more perceivable and do not designate a quality or a property of a person or a thing. When they are nominalized they tend to be more objective, more scientific and less specific than imperceptible verbs, which, in contrast, are less objective in terms of the perceived concepts, less scientific, more personal and more specific. Also, nominals with khwaam- tend to signify an amount or volume of a certain quality, which is mostly value-oriented. In addition, imperceptible verbs are normally adjectival verbs, which equal adjectives in English.
Nominalization and categorization of verbs in Thai

References


NEGATION IN KAVALAN

Li-May Sung and Yu-Ting Yeh
Graduate Institute of Linguistics
National Taiwan University
limay@ntu.edu.tw; r91142001@ntu.edu.tw

1. Introduction
One issue in the studies of negation concerns categorizing the negative elements (as in Dahl 1979; Payne 1985; Ouhalla 1990; Kahrel and van den Berg 1994; Haegeman 1995; Mosel 1999 and others). Identifying the syntactic category of a negative element not only explains its syntactic distribution, but also helps to discuss the syntactic structure of the corresponding negative construction.

This paper attempts to examine three hypotheses of the syntactic category of the negator mai in Kavalan. We will argue for analyzing mai as a negative auxiliary verb.

2. Literature Review
Mosel (1999) mentions that negative elements may belong to a wide range of syntactic categories, such as verbs, auxiliaries, particles, affixes, and so on. Payne (1985) also proposes some criteria for recognizing a negative verb. It is claimed that “a negative verb always has at least some properties of regular verbs, such as occurrence with a verbal complementizer or inflection for mood, tense, aspect, person, or number” (Payne 1985: 207). Payne further divides negative verbs into two major groups: one is a higher verb taking a full sentential complement and the other is “a finite auxiliary verb to the lexical verb, which in turn typically occurs in some non-finite form” (Payne ibid: 207).

According to Payne (1985), for a negative marker to be analyzed as a higher verb, two features are required. One is the existence of a sentential boundary between the negative verb and the lexical verb. Take Fijian for example, there is a complementizer ni between the negative verb sega and the full sentential complement (as shown in (1)).

(1) Fijian (Payne 1985: 210)
E sega [ni a yacomai ena siga Vakaraubuka ko Jone]
PCL Neg that Past arrive on day Friday ART John
‘John didn’t arrive on Friday.’

The other feature is that the complement selected by the higher negative verb is generally finite in the sense that it may contain the tense/aspect markings and/or the subject. For example in Tongan, the higher negative verb ikai takes a complement

---

1 Research represented in this work was financially supported by the National Science Council (NSC93-2411-H-002-094), Taiwan, granted to Dr. Li-May Sung. We would like to thank our Kavalan speakers Buya, Imui, and Abas for their constant assistance and patience.

2 This paper focuses on the most frequently used negator mai. There are other negative elements in Kavalan, such as usa ‘be not’, naRin ‘don’t’, sukaw ‘not good; not allowed’; Rayngu ‘not know; incapable of’ and taga ‘not want’ (see Yeh (2005)).
that obligatorily contains an embedded aspect *ke* and the subject pronoun *ne*, as shown in (2).

(2) Tongan (Payne 1985: 209)

a. *Na’e ikai [s ke ne fai ‘a e ngauue]*
   
   Asp Neg Asp he do Abs the work
   ‘He didn’t do the work.’

b. *Na’a ne ikai [s ke fai ‘a e ngauue]*
   
   Asp he Neg Asp do Abs the work

As seen in (2b), moving the subject pronoun *ne* out of the embedded complement to the matrix clause results in ungrammaticality.

On the other hand, Payne proposes that, in the purest case, a negator being classified as a negative auxiliary verb is “marked with all the basic verbal categories of person, number, tense/aspect and mood (if these are realized in the language concerned), whereas the lexical verb assumes an invariant, participial form. [Moreover,] there will be no evidence, like the presence of complementizers, for a full sentential boundary between the negative verb and the lexical verb…” (Payne 1985: 212). Evenki negator *ə* is such a pure auxiliary verb, as illustrated in (3).

(3) Evenki³ (Payne 1985: 213)

a. *Nuŋaŋ baka-ŋkī-n*
   
   he find-Past-3Sg
   ‘He found.’

b. *Nuŋaŋ ə-ŋkī-n baka-ra*
   
   he Neg-Past-3Sg find-PART
   ‘He didn’t find.’

In the affirmative sentence (3a), the tense inflection -ŋkī and the person/number agreement are carried by the lexical verb. In the negative counterpart (3b), however, both of these inflections are carried by the negative auxiliary verb *ə*, whereas the lexical verb *baka* takes a participial form suffixed by -ra instead. In addition, as seen in (3b), there is no complementizer or other evidence for a full sentential boundary between the negative auxiliary *ə* and the following lexical verb.

*mai*, the most frequently used negator in Kavalan which is a verb-initial language, appears on initial examination to behave like a verb, since it occurs clause-initially, attracts nominative bound pronouns and carries tense/aspect markers as shown in (4)-(6)⁴.

---

³ Evenki is “a member of the northern or Siberian subgroup of the Tungus family” (Payne 1985: 212).

⁴ Glossing: AV: Actor Voice; NAV: Non-actor Voice; Ncm: noun-class marker; Pfv: perfective; DM: discourse marker; Neg: negator; Nom: nominative; Obl: oblique; Gen: genitive; Sg: singular; Compl: complementizer; Loc: Locative.
Negation in Kavalan

(4) (Occur in clause-initial position)

a. \textit{p-m-ukun tu wasu 'nay ci buya}
\textit{AV-hit Obl dog that Ncm BUYA}
‘Buya hits that dog.’

b. \textit{mai p-m-ukun tu wasu 'nay ci buya}
\textit{Neg AV-hit Obl dog that Ncm BUYA}
‘Buya doesn’t hit that dog.’

(5) (Attract bound pronouns)

a. \textit{p-m-ukun-iku tu wasu 'nay}
\textit{AV-hit-1Sg.Nom Obl dog that}
‘I hit that dog.’

b. \textit{mai-iku p-m-ukun tu wasu 'nay}
\textit{Neg-1Sg.Nom AV-hit Obl dog that}
‘I don’t hit that dog.’

(6) (Carry tense/aspect markers)

a. \textit{p-m-ukun-ti tu wasu 'nay ci buya}
\textit{AV-hit-Pfv Obl dog that Ncm BUYA}
‘Buya have hit that dog.’

b. \textit{mai-ti p-m-ukun tu wasu 'nay ci buya}
\textit{Neg-Pfv AV-hit Obl dog that Ncm BUYA}
‘Buya doesn’t hit that dog anymore.’

This paper therefore attempts to examine the syntactic category of \textit{mai} based on Payne’s (1985) proposals. In addition to testing the possibilities of \textit{mai} being a higher verb or an auxiliary verb, we also consider the third possibility: \textit{mai} as an initial main verb (V1) taking a verbal complement as in a serial verb construction. We conclude that \textit{mai} is a negative auxiliary.

The presentation is organized as follow. We at first briefly introduce Kavalan, and then present the examinations of \textit{mai} as an initial main verb in a serial verb construction, a higher verb taking a full sentential complement, or an auxiliary verb to lexical verb. Finally we will give a conclusion.

3. Description of Kavalan

Kavalan is one of the Formosan languages spoken by the aboriginals living in the plains of Eastern Taiwan. The population is about 820 now\footnote{The data is cited from the website of the Council of Indigenous Peoples, Executive Yuan, Taiwan. (http://others.apc.gov.tw/popu/9403/aprp5803.htm)}; however, less than 100 can speak Kavalan fluently (Chang 2000a). In Nov. 2002, Kavalan tribe is legitimately recognized as the 11th Taiwan aboriginal tribe.

Kavalan is a verb-initial language, and the basic word order is VOS in that the grammatical subject in nominative case tends to be placed in the sentence-final position.
As can be seen, Kavalan has two voice options, Actor Voice (AV) and Non-actor Voice (NAV). In AV clauses, the actor is selected to be the sentence subject (as in (7)), while in NAV clauses, it is the non-actor (e.g., patient) that is the grammatical subject (as in (8)). The actor in a NAV clause is marked with genitive case (8), and the patient in an AV clause is marked with oblique case (7).

Moreover, there are two types of bound pronouns in Kavalan: nominative bound pronouns (such as –iku in (5)) and genitive bound pronouns (such as –na in (8)). Chang (1997; 2000a) has argued that these two types of pronouns have different morphological status. The nominative ones are pronominal clitics, while the genitive ones are verbal agreement affixes that only occur in NAV clauses.

In terms of negation, mai is the most commonly used negator in Kavalan. It is used to negate sentences with all types of non-nominal predicates (including verbal, adjectival, and locative, as in (9)-(11)). The adding of mai to the clause-initial position of the positive counterparts does not change the word order or the form of the main predicates.

(9) (Verbal)
   a. \( p-m-ukun\) \( tu\) \( wasu\) ‘nay \( ci\) \( buya\)
      AV-hit Obl dog that Ncm BUYA
   ‘Buya hits that dog.’

   b. mai \( p-m-ukun\) \( tu\) \( wasu\) ‘nay \( ci\) \( buya\)
      Neg AV-hit Obl dog that Ncm BUYA
   ‘Buya doesn’t hit that dog.’

(10) (Adjectival)
    a. \( ibabaw\) \( ci\) \( buya\)
       tall Ncm BUYA
    ‘Buya is tall.’

---

6 Based on discourse data, Yeh (2004) argues for a strong tendency of VOS order in Kavalan, contrary to the common observation of word order as VSO in AV clauses as discussed in Lee (1997), Chang (1997, 2000a) and Liao (2004).

7 In Yeh et al. (1998), negative locative sentences and negative possessive/existential sentences are grouped together. Kavalan is however in a different case. Negative locative sentences, structurally different from negative possessive/existential sentences, are rather parallel to, and therefore should be classified with negative declarative constructions.
Negation in Kavalan

b. **mai** ibabaw ci buya
   Neg  tall   Ncm  BUYA
   ‘Buya is not tall.’

(11) (Locative)
   a. ta-nawung-an ci abas
      Loc-mountain-Loc Ncm  ABAS
      ‘Abas is in the mountain.’

   b. **mai** ta-nawung-an ci abas
      Neg Loc-mountain-Loc Ncm  ABAS
      ‘Abas is not in the mountain.’

4. Syntactic Category of **mai**

We now turn to examine the syntactic category of **mai**. As has been mentioned, **mai** appears to act like a verb due to its clause-initial position and its ability to attract nominative bound pronouns as well as tense/aspect markers (as in (4)-(6)). There are three possibilities of the syntactic category of **mai**. First, **mai** may be the initial main verb (V1) taking a verbal complement as in a serial verb construction. Second, **mai** may be a higher negative verb selecting a full (finite) sentential complement, literally equivalent to ‘It is not (true) that ….’. And third, **mai** may act as a finite auxiliary to the following lexical verb. These three possibilities are examined respectively.

4.1 **mai** as V1 in serial verb constructions

We at first test if **mai** functions as a main verb taking a verbal complement as in a serial verb construction. In some respects, Kavalan negative declarative constructions seem to be similar to serial verb constructions, as proposed in Lord (1993), Stewart (2001), and Crowley (2002), where two (or more) verbs share a single subject without a covert connective marker in between. For instance, neither the complementizer *tu* nor the conjunctive *atu* ‘and’ can occur between two verbs in a serial verb constructions (as in (12)), or between **mai** and the following verb (as in (13)).

(12) paska *tu/*atu q-em-an tu Rak ci abas
    try *Compl/*and AV-eat Obl alcohol Ncm  ABAS
    ‘Abas tries to drink alcohol.’

(13) **mai** *tu/*atu q-em-an tu Rak ci abas
    Neg *Compl/*and AV-eat Obl alcohol Ncm  ABAS
    ‘Abas doesn’t drink alcohol.’

Moreover, tense/aspect markers obligatorily attach to **mai** instead of to the following lexical verb in a negative declarative construction (as in (15)), as they do to V1 in a serial verb construction (as in (14)).
Li-May Sung and Yu-Ting Yeh

(14) a. paska-ti q-em-an tu Rak ci abas
   try-Pfv AV-eat Obl alcohol Ncm ABAS
   ‘Abas has tried to drink alcohol.’
   b. * paska q-em-an-ti tu Rak ci abas

(15) a. mai-ti q-em-an tu Rak ci abas
   Neg-Pfv AV-eat Obl alcohol Ncm ABAS
   ‘Abas doesn’t drink alcohol anymore.’
   b. * mai q-em-an-ti tu Rak ci abas

However, treating negative declarative sentences as serial verb constructions in which mai functions as V1 would face two problems. First, the initial verb (V1) in a serial verb construction must be a lexical verb that names an action, event, process, or state (Sebba 1987; Lord 1993; Stewart 2001). mai in (13) and (15) is however not so semantically contentive as paska ‘try’ in (12) and (14); it merely conveys a negative value to the following lexical verb ‘eat’.

A more serious problem lies in that the non-initial verb (V2) in a true serial verb construction in Kavalan is strictly restricted to be in AV form, while V2 (i.e., the lexical verb following mai) in a negative declarative construction can be in either AV or NAV form. Consider the examples (16) and (17).

(16) a. paska-iku q-em-an tu Rak
   try-1Sg.Nom AV-eat Obl alcohol
   ‘I try to drink alcohol.’
   b. * paska qan-an-ku tu/(ya) Rak
      try eat-NAV-1Sg.Gen Obl/(Nom) alcohol
   c. paska-an-ku q-em-an (ya)/*tu Rak ‘nay
      try-NAV-1Sg.Gen AV-eat (Nom)/*Obl alcohol that
      ‘I try to drink the alcohol.’
   d. * paska-an-ku qan-an (ya) Rak ‘nay
      try-NAV-1Sg.Gen eat-NAV (Nom) alcohol that

(17) a. mai-iku q-em-an tu Rak
   Neg-1Sg.Nom AV-eat Obl alcohol
   ‘I don’t drink alcohol.’
   b. maiqan-an-ku (ya)/tu Rak ‘nay
      Neg eat-NAV-1Sg.Gen (Nom)/Obl alcohol that
      ‘I don’t drink the alcohol.’

As can be seen, no matter whether V1 in a true serial verb construction is in AV (as in 16a, b) or NAV (as in 16c, d) form, V2 must be in AV form. But V2 (i.e., the verb following mai) in a negative construction can be freely in AV (as in 17a) or NAV (as in 17b) form. In addition, in a true serial verb construction, the case of the theme
Negation in Kavalan

argument is assigned by V1 instead of by V2. Therefore, Rak ‘alcohol’ in (16c) takes nominative case (from *paska-an*) instead of oblique case (from *q-em-an*). By contrast, the theme argument of a negative declarative construction gets its case from V2 (i.e., the verb following *mai*). As shown in (17b), Rak ‘alcohol’ takes nominative case (from *qan-an*) rather than oblique case.

In summary, the evidence above stands against analyzing *mai* as the initial main verb (V1) taking a verbal complement as in a serial verb construction.

4.2 *mai* as a higher verb

Next we examine whether *mai* is a higher verb taking a full sentential complement, literally equivalent to ‘It is not true that…’. According to Payne (1985), two features are required for a negative marker to be classified as a higher verb. First, there should be a full sentential boundary, such as a complementizer, between the negative verb and the lexical verb. And second, the sentential complement selected by the higher negative verb is generally finite and contains its own tense/aspect markings and/or the subject.

Kavalan *sazmaken* ‘believe’ provides an example of a higher verb taking a finite sentential complement. Consider the examples (18)-(20).

(18) *sazmaken-iku* (*tu*) *q-em-an-ti* *ci* *buya* *tu* *Rak*

believe-1Sg.Nom (Compl) AV-eat-Pfv Ncm BUYA Obl alcohol

‘I believe (that) Buya has drunk alcohol.’

(19) *sazmaken-ti-iku* (*tu*) *q-em-an* *ci* *buya* *tu* *Rak*

believe-Pfv-1Sg.Nom (Compl) AV-eat Ncm BUYA Obl alcohol

‘I have believed (that) Buya drinks alcohol.’

(20) *sazmaken-ti-iku* (*tu*) *q-em-an-ti* *ci* *buya* *tu* *Rak*

believe-Pfv-1Sg.Nom (Compl) AV-eat-Pfv Ncm BUYA Obl alcohol

‘I have believed (that) Buya has drunk alcohol.’

The optional complementizer *tu* in (18)-(20) indicates a sentential boundary between *sazmaken* ‘believe’ as a matrix higher verb and its clausal complement. Moreover, example (18) proves the embedded clause to be finite since it contains its own aspect -ti that scopes only over the embedded domain, not over the matrix domain. The aspectual marker can of course attach to the matrix verb *sazmaken* as in (19); but in this case, the aspectual scope is restricted to the matrix clause. Example (20) demonstrates that the matrix and the embedded clauses can have their own aspects at the same time.

We argue against treating *mai* as a higher verb as *sazmaken* selecting a full sentential complement because negative declarative constructions in Kavalan lack the two features mentioned above. The unacceptability of the occurrence of a complementizer *tu* between *mai* and the following lexical verb is attested in (21).

(21) *mai* *tu* *q-em-an* *tu* *Rak* *ci* *abas*

Neg *Compl AV-eat Obl alcohol Ncm ABAS

‘Abas doesn’t drink alcohol.’
Most crucially, when there is a perfective marker -ti or a future marker -pa, it obligatorily attaches to mai rather than to the following lexical verb, as shown in (22) and (23) below.

(22) a. **mai-ti q-em-an tu Rak ci abas**
    Neg-Pfv AV-eat Obl alcohol Ncm ABAS
    ‘Abas doesn’t drink alcohol anymore.’

    b. * **mai q-em-an-ti tu Rak ci abas**

    c. * **mai-ti q-em-an-ti tu Rak ci abas**

(23) a. **mai-pa q-em-an tu Rak ci abas**
    Neg-Fut AV-eat Obl alcohol Ncm ABAS
    ‘Abas will not drink alcohol.’

    b. * **mai qan-pa tu Rak ci abas**

    c. * **mai-pa qan-pa tu Rak ci abas**

The ungrammaticality of (22b) and (22c) as well as of (23b) and (23c) demonstrates that mai behaves differently from sazmaken, which is a true higher (matrix) verb taking a full sentential complement. If the constituents following mai were indeed a sentential complement, the tense/aspect markers should be possible to attach to the verbal predicate of the complement, contrary to the fact.

Furthermore, the placement of nominative bound pronouns in negative declarative constructions provides additional evidence against analyzing the constituents following mai as finite sentential complement.

Consider example (24).

(24) a. **mai q-em-an-iku tu Rak**
    Neg AV-eat-1Sg.Nom Obl alcohol
    ‘I don’t drink alcohol.’

    b. **mai-iku q-em-an tu Rak**
    ‘I don’t drink alcohol.’

This example demonstrates clitic climbing which is a type of head movement that will be blocked by intervening C, finite Infl, Neg, and clitics (Rosen 1989; Moore 1994; Haegeman 1995; Chang 1997). The fact that the nominative pronominal clitic -iku in (24) can undergo clitic climbing suggests that there is not an intervening T’-head between mai and the following lexical verb. If the constituents following mai were a sentential complement, there would be a finite Infl or C projection which would block clitic climbing, contrary to the fact.

Evidence above proves that mai is not a higher negative verb taking a full sentential complement.

4.3 mai as an auxiliary verb

Finally, we consider the possibility of mai being a negative auxiliary verb to the lexical verb. Previous discussions have revealed that mai in Kavalan behaves in many ways like an auxiliary verb. For one thing, the existence of the complementizer tu as a full
Negation in Kavalan

sentential boundary between *mai* and the following lexical verb is disallowed, as shown in (21).

For the other, the tense/aspect markers -*ti* and -*pa* that originally attach to the lexical verbs in affirmative declarative constructions are obligatorily attracted to *mai* in negative counterparts, as shown in (25) and (26)

(25) a. *q-em-an-ti* tu Rak ci abas  
    AV-eat-Pfv Obl alcohol Ncm ABAS  
    ‘Abas has drunk alcohol.’

b. (i) *mai-ti* *q-em-an* tu Rak ci abas  
    Neg-Pfv AV-eat Obl alcohol Ncm ABAS  
    ‘Abas doesn’t drink alcohol anymore.’

(ii) * mai q-em-an-ti* tu Rak ci abas

(26) a. *qan-pa* tu Rak ci abas  
    eat-Fut Obl alcohol Ncm ABAS  
    ‘Abas will drink alcohol.’

b. (i) *mai-pa* *q-em-an* tu Rak ci abas  
    Neg-Fut AV-eat Obl alcohol Ncm ABAS  
    ‘Abas will not drink alcohol.’

(ii) * mai qan-pa* tu Rak ci abas

In addition to tense/aspect markings, *mai* serves as well as an obligatory carrier of the suffix -*pama* ‘yet’ and of the hedge discourse marker -*ma*, as shown in (27) and (28) respectively.

(27) a. *q-em-an-pama* tu Rak ci abas  
    AV-eat-still Obl alcohol Ncm ABAS  
    ‘Abas still drinks alcohol.’

b. (i) *mai-pama* *q-em-an* tu Rak ci abas  
    Neg-yet AV-eat Obl alcohol Ncm ABAS  
    ‘Abas hasn’t drunk alcohol yet.’

(ii) * mai q-em-an-pama* tu Rak ci abas

(28) a. *q-em-an-ma* tu Rak ci abas  
    AV-eat-DM Obl alcohol Ncm ABAS  
    ‘Abas drinks a little alcohol.’

b. (i) *mai-ma* *q-em-an* tu Rak ci abas  
    Neg-DM AV-eat Obl alcohol Ncm ABAS  
    ‘Abas doesn’t drink much alcohol.’

(ii) * mai q-em-an-ma* tu Rak ci abas

In view of the evidence above, we would like to propose that *mai* in Kavalan
negative declarative constructions is more like a negative auxiliary verb. Nevertheless, it is notable that *mai* is not so pure as the case in Evenki (as in (3)) where the negative auxiliary attracts all the verbal inflections, including person/number, tense, aspect and mood. Payne (1985) mentions it is possible that person and/or number distinctions may totally fail to be manifested on the negative auxiliary verb, with the result of compensatory markings on the lexical verb. This is also the case concerning Kavalan genitive bound pronouns which have been argued in Chang (1997) to be verbal agreement affixes, as shown in (29).

(29) a. mai pukun-an-na, ni buya, ci abas
   Neg hit-NAV-3Sg.Gen Gen BUYA Ncm ABAS
   ‘Buya doesn’t hit Abas.’

b. *mai-na, pukun-an ni buya, ci abas

As can be seen, in negative declarative constructions with NAV lexical verbs, the genitive bound pronouns must attach to the NAV lexical verb (as in (29a)) rather than to the negative auxiliary verb *mai* (as in (29b)). This suggests that *mai* is not a purest auxiliary that can attract all the verbal inflections. If it were, we would find it possible to carry the person/number agreement affixes as the tense auxiliaries *wada* and *maha* do in Seediq, another Formosan language, as illustrated in (30).

(30) Seediq (Chang 1997: 99)
a. wada-*ku-na*, bube-un na pawan, ka yaku
   Past-1Sg.Nom-3Sg.Gen beat-NAV Gen PAWAN Nom 1Sg
   ‘Pawan beat me.’

b. maha-*ku-na*, bube-un na pawan, ka yaku
   Fut-1Sg.Nom-3Sg.Gen beat-NAV Gen PAWAN Nom 1Sg
   ‘Pawan will beat me.’

In (30), tense auxiliaries in Seediq, unlike the negative auxiliary *mai* in Kavalan, can carry the person/number agreement affixes (i.e., *-ku* and *-na* in this case).

5. Conclusion and implications

In conclusion, we claim that *mai* in Kavalan is not an initial main verb taking a verbal complement as in a serial verb construction or a higher verb selecting a full sentential complement. Instead, *mai* acts as a negative auxiliary verb that attracts most of the verbal inflections except for person/number agreement affixes.

The finding in this study helps to further examine the structural position of *mai* within the framework of Principles-and-Parameters Theory (also known as Government and Binding Theory). Based on the finding, Yeh (2005) claims that *mai* heads a functional negative projection (NegP) between TP and AgrP and functions as a potential intervening head which blocks V-movement (see also Pollock 1989; Ouhalla 1990; Belletti 1990; Haegeeman 1995). This phrase structure explains the syntactic distribution of verbal inflections (such as tense/aspect and person/number agreement affixes) in Kavalan negative declarative constructions.

In addition to the declarative constructions as discussed in this paper, *mai* can
Negation in Kavalan

also appear in possessive/existential constructions as well. In this respect, Kavalan differs typologically from many other Formosan languages (Amis, Atayal, Bunun, Paiwan, Puyuma, Rukai, Seediq, Tsou, Yami), which use different negators in possessive/existential and declarative constructions, in that its negative possessive/existential are marked by the same negator *mai* though *mai* in these two types of constructions belongs to different syntactic categories. In possessive/existential constructions, *mai* functions as a lexical main verb rather than as a negative auxiliary verb. As shown in (31)-(32), *mai* is the sole verbal predicate and in terms of the case realization of nominal arguments, *mai* in possessive/existential constructions behaves as a (negative) transitive verb which obligatorily assigns only oblique case to the possessed or existed entity.\(^8\) This observation accords with Zeitoun et al. (1999: 40) in which they suggest that in Formosan languages “the constituent usually heading existential, possessive, and locative constructions should be treated as a verb”.\(^9\)

(31) *mai* in possessive constructions

a. *mai *(tu) wasu ci buya
   Neg Obl dog Ncm BUYA
   ‘Buya doesn’t have a dog.’

b. *mai-*iku *(tu) kelisyu
   Neg-1Sg.Nom Obl money
   ‘I don’t have money.’

(32) *mai* in existential constructions

a. *mai* tu/*ya) benina (*’nay) ta babaw na takan
   Neg Obl/*Nom) banana *that Loc above Gen table
   ‘There is no banana on the table.’

b. *mai* tu/*ya) razat (*’nay) ta-repaw-an
   Neg Obl/*Nom) person *that Loc-house-Loc
   ‘There is nobody home.’

---

\(^8\) In possessive constructions, if the possessed is replaced with a nominative case, the meaning will change and the implication of possession does not exist anymore.

(i) *mai* *(ya)/tu wasu ni buya
   Neg (Nom)/Obl dog Gen BUYA
   ‘Buya’s dog is gone/dead.’
   * ‘Buya doesn’t have a dog.’

(ii) *mai* *(ya)/tu kelisyu-ku
   Neg (Nom)/Obl money-1Sg.Gen
   ‘My money is gone (lost).’
   * ‘I don’t have money.’

\(^9\) This claim is only partially true for the Kavalan data because *mai* in possessive/existential constructions indeed functions as a verb, *mai* in locative sentences is syntactically parallel to negative declarative constructions and thus should be treated as a negative auxiliary rather than as a main verb.
In addition to a preliminary descriptive study in Chang (2000a), this paper provides a categorical analysis of the negator *mai* in Kavalan. Moreover, the present study also provides a valuable contribution to Formosan typological studies in addition to Yeh et al. (1998) in which negative constructions of Amis, Atayal, Paiwan, Rukai, Saisiyat, and Tsou are examined.

References


Today there are nearly three million Vietnamese living in more than 90 countries and territories the world over. However, this does not mean that these people speak all the languages of the countries of their residence. Like any other migrant community, the Vietnamese diasporas maintain their culture via their mother tongue. Due to different language ecology factors, the Vietnamese language used by these people can be called “a migrant Vietnamese”. The paper aims to give out initial observations made on the nature of phonological, lexical and structural borrowings, integration as well as transferences of the Vietnamese used in Australia.

1. Historical background
The existence of the Vietnamese overseas can be dated back, firstly, from early the 8th Century, when Prince Ly Long Tuong, the second son of King Ly Anh Tong, took refuge in Korea or later when the Ming dynasty invaded Vietnam and brought back with them a number of monks, and workers to China (Tran, p. 19-21). In Japan, early in the Seventeenth Century, Princess Ngoc Van of the Nguyen Dynasty, called Anio, was married to Araki Sotaro and lived in Nagasaki since then. In the Eighteenth Century, a number of Catholic disciples from Vietnam migrated to Thailand. From the second half of the Nineteenth Century to 1954, a number of Vietnamese immigrants arrived in France and other French colonies Tahiti and New Zealand. However, it could be said that a remarkable number of Vietnamese immigrants was seen in Southeast Asian countries in late the Nineteenth and early in Twentieth Centuries, the most noteworthy was from such Patriotic Movements headed by Phan Boi Chau (in Japan, and China in 1904), Tang Ban Ho (in Japan, China, Thailand 1904-1911), Pham Hong Thai (in China, 1918). This could be said as the second turning point of Vietnamese overseas immigration. The third turning point was that, a number of youngsters who sought ways for national salvation by going to study overseas. Typical of this trend was Phan Chu Trinh (in France 1911-1925), King Ham Nghi (Algeria, 1888-1947), both King Thanh Thai (1915-1947), King Duy Tan (1916-1945) in La Reunion, Africa. However, the formation of the Vietnamese diaspora was resulted from such events in between 1940s-1950s during the Second Indochina War (Tran, 1997 & Carruthers, 2004) in Laos, Thailand, Cambodia, France, Japan, and the American War (1954-1975) in Western nations, USA, Australia, France, Canada and in some Eastern Bloc countries. Statistically speaking, the number of Vietnam-born in the US, Australian and Canada are the most significant: 1,122,528 (2000 census), 154, 830 (2001 census (136,810 (1996 census) in the US, Australia, and Canada respectively.

Numerous studies have investigated the factors and institutions promoting the maintenance of community language other than English (CLOTES) (Clyne, 1967, 1970; Haugen 1971, 1979; Clyne 1967, 1985; Klarger, 1976, Smolicz and Harris 1976) stressing on the wholesomeness for the human environment of the maintenance and
development of multilingualism. According to sociologists (Fishman 1965; Cooper 1969; and Greenfield 1970), the domain of code selection of bilinguals and multilinguals depends largely on interlocutors and such variables as ethnic ascription, situation of speaking, topic, style, role-relationship, venue, interaction type and medium (Sandkoff 1971). While in many families, English foreigner talk replaces the ethnic language as the second generation children’s code of communication (Clyne 1985, p. 58), CLOTEs have been used as a main vehicle for maintaining cultures, and the sole means of communication for the first generation as well as for the monolingual groups in most urban diasporas. In their turns, these ethnic languages used in these diasporas bear quite a lot of linguistic borrowings (Haugen 1950; Myers-Scotton 1977, 1988; Poplack and Vanniaraajan 1990 and Myers-Scotton and Jake 2000). In Australia, as pointed out by Clyne (1985: 94), there are as many varieties of “migrant languages” as there are speakers, since the nature and degree of English influence and general adaptation of the base language to the Australian context will largely depend on the individual speakers’ activities and the lifestyle as well as his or her experience in both languages. This study presents the characteristics of the Vietnamese being used in Australia culled from more the conversations of Vietnamese/ Australian bilinguals and monolinguals in Sydney, Melbourne and ACT and a number of prints available within the Vietnamese communities.

2. Methodology
The data collected in this study was from 37 conversations recorded on 26 Vietnamese adults (of 1.5 generation) aged between 22- 62, 17 males, and 9 females. They are of different professions, IT engineers, dentists, doctors, professors, writers, catering services, take-away servers, bakers and housewives etc. Among the 26 speakers, only 4 of them were fluent bilinguals. All the participants have at least 5 years living in Australia.

The topics of the conversations are all about daily activities and ways of life. All the interactions before and after the recordings were in Vietnamese. The conversations could be described as free-flowing descriptions, discussions or culture-specific craft explanations. In order to minimize the social distance between the field worker and the informants, I took part in the conversation from time to time. The geographical areas of data collection were Mt. Pritchard (NSW), Springvale (VIC) and Belconnen (ACT). For each data, I transcribed all the instances of thee substantives uttered which includes nouns, adjectives, verbs, interjections or any item which functions as significant linguistic units. Besides, about more than 50 texts, including short stories, articles, news of various lengths were investigated. Both lexical and structural traits were major points of focus.

3. Linguistic analysis
3.1. Forms of lexical renewal
According to Clyne, three common ways that speakers shape their vocabulary to meeting the changing needs of life (Clyne 1985: 94) are neologism, semantic expansion and transference. In this case, neologisms in the Vietnamese created in Australia are based on the existing morphological devices of the language: Eg.
Preliminary observations on a “migrant Vietnamese”

dì làm farm
tách form/ ghép form
ăn welfare, ăn ticket
làm nail
làm Centerlink
di shop, sang shop
xin council

In Vietnamese, di làm ruộng, or làm nông (working on the farm) refers farming work in general, which involves ploughing, watering, harvesting and most of the work is manual. Di làm farm refers to fruit-picking on seasonal base. Sometimes it refers to a cash-paid seasonal labour. Di shop does not, however, mean “going shopping”, but refers to “going for a specific purpose, as in di sop tâu, di sop thịt, sop cá, or shop trái cây etc. whereas in English, sop Tâu refers to Asian Groceries, sop thịt butchers’, sop cá fish shop, sop trái cây fruit shop. The word “sop” (shop) is actually expanded semantically.

Meanwhile, neo (nail) as in làm neo, tiêm neo, tí tết or tí tật (ticket) as in ăn tí tật (got fines), Lêm (claim) as in lêm thuế/ bảo hiểm (tax/ insurance claim) are examples of semantic expansion due to the influence of English homophones or the preference of archaism in the migrant language (Clyne 1985). Contrary to this tendency, lexical and semantic changes in Vietnamese are found resisted against, and in some papers, or conversations, some words (no longer used in Vietnam) are being used within the diasporas, especially, in adult groups (of over 50), which forms a stabilized use of archaisms.

Trong sáng thứ Tư tuần qua, thủ lãnh (meaning: leader, current use: lãnh đạo) Đảng Lao Động Kim Beazley lên tiếng chỉ trích …
(Last Wednesday morning, Leader of The Labour Party Kim Beazley raised his criticism …)

(From Tự Do Không Tranh Cử, in Nam Úc Thời Báo online)

Khởi di (meaning: beginning, current use: bắt đầu) từ sự thay thuận này, môn tiếng Anh là môn học đầu tiên mà …
(From this agreement, English is the first subject that …)

(From Trường học trên toàn nước Úc có chương trình học thường nhất, in Thời Báo, issue 300, 21/7/2003)

Sang Thương Vụ (meaning: business; current use: doanh nghiệp or cxa sô làm ăn) : Ứí đáp. Có hội làm ăn rất tốt, không cạnh tranh, làm nhiều ăn nhiều… Cần tiền sang gấp. Xin liên lạc…

(Thời Báo-Vietnamese Community Newspaper, issue 300, 21/7/2003)
Bao Duy Thai

Ở bên Việt Nam ai có... báo giấy bằng nhựt trình (meaning: newspapers, current use: báo)

(Tư Éch Di Dân, Việt Luận Online, issue: 18/2/05)

Nhà chức trách Úc đã yêu cầu sự giúp đỡ của công chúng ở Đức trong một cô gái từng ra cănカード (meaning: identity card, current use: giấy chứng minh/chứng minh nhân dân) của Phụ nữ này.

(Một Phụ Nữ Sydney Bị Giảm Giữ Nhầm Lẫn Trong Trại Di Trú, Nam Úc Thời Báo, 18/2/05)

Though maintaining archaisms is most seen in the names of countries, such as, Hoa Lục, Nam Dương, Nhật Bồn, Mạc Tứ Khoa, Luận Dọn, Hoa Thịnh Đốn, Áng- Lê, Úc Đại Lợi, the corpus also shows archaisms appear in various parts of speech: nouns, verbs, adjectives, adverbs etc. These words are rarely found in the diasporas due to lack of experiences or member of communication networks, especially those in home and business domains indicating daily routines. They are: khăn trường, phân khối, hồ hối, mạnh dân, nhân thức, đăng ký, quán chúng, báo cập, hồ khấu etc...

3.2 Forms of lexical transference

The most usual mean, as pointed out by Clyne, in “migrant languages” is through transference in English, through the idiolects of the first generation migrants in the urban “melting pot situation”(Clyne 1985). It affects practically all speakers, both bilinguals and monolinguals in various extents. Strangely, from the recordings, we found that lexical transference is not necessarily concomitant with a higher proficiency in English, but in many ways, it can be attributed to the differences in the lifestyles and preoccupations in the country of origin. Categories of lexical transference could be as follows.

a. Contextual: words particularly prevalent in English references to the work, occupations, school or home domains: shop thíết, shop hoa, shop vải, shop Tàu (shop thực phẩm ở châu), shop nail, làm ga-dơn (garden), a-kao-tờn (accountant), xen-tờ-lin (Centerlink), thiết- ở quê (take-away), lin nhà (cleaning), thằng boy (boy), con go (girl), ê-dön (real estate agent), lót- kẻ (locker), goa ropp (wardrobe), bò rít vở nia (brick veneer), dì bòt (going by bus), lấy hở lì dế (take holidays), bấy chéc (pay check), bí-zí-nít (business), go rần tì (warranty), in-sua rần (insurance), công trếc (contract), phom (form) (diện form, khai form), sít tem tiêu- ti (stamp duty), kao sô (council), ờt stó-dí (Aus Study), dai ờt (diet), phó lết or lét (flat).

b. While nouns are the most common word class transferred (Haugen 1953: 406; Clyne 1985: 95) due to the direct link between form and content, other of parts of speech are as follows:

Verb: lin (clean), bút (book), thiết (take), rán (run), mê nịt (manage), hen-dờ (handle), điều (deal), thiết ờ vò (take over), li (lease), seo (sell), ruy lết (relax), lút áp tờ (look after), thiết ke (take care), ke (care), nó gọ shi ệt (negotiate), rin (ring/ call), ko (call), phon (phone), ken sô (cancel), guốc pát
Preliminary observations on a “migrant Vietnamese”

Thai (work part time)/ phun thai (full time), ga răn ti (guarantee), pho gét (forget), it no (ignore), phải nen (finance)

Adjectives/Adverbs: i-zì (easy), bi-zì (busy), hép pi (happy), lật kì (lucky), te-ri-bồ (terrible), sốc (shocked), nai (nice), ke-fun (careful), xế (safe), gút (good), phờ/phi (free)

E.g., Con gớ đồ lạt kì ghể. Xấu mà sống hệp pi lắm nghe. Được thằng bố thiệt nai
(That girl is lucky indeed. She’s not good-looking, but got a happy life. She’s got a real nice boyfriend)

(Conversation recorded)

Pronouns: du (you), mi (me/I), and personal pronouns and adjective pronouns are used interchangeably.

E.g., Du lấy xe của mi mà đi làm
(You can use my car to get to work)

(Conversation recorded)

Or in the following conversation of the second-generation kids on a commercial VCD available at a local shop:

- Cái nay me tăng cho you nè. Me giũ một cái, you giũ một cái. Chừng nào you nhìn thấy nò, you nghĩ tôi me. (Well, I give you this. You keep one, and I keep one. Whenever you see it, you would think of me.)
- A, you lucky hơn me nhiều lắm: vi you có ba nè, you có mà nè. Mỗi lần me đi học về, me buồn lắm. Me khóc hoài à...Er. Má me nói me là...hồng có me, là mà me kills herself. Rồi ba me nói me là Ba me thường me với Má me nhiều nhất trên đời. (You’re much luckier than me, ‘cause you’ve got your Dad and Mom. I cry all the times. Eh, Mom told me that without me she would have killed herself. And Dad told me that he loves me and Mom best of all in the world.)
- Oh, oh, nhưng mà Ba you xào quá à. Néo mà, Ba you thường you với Má you luôn hà, thì Ba you đâu có di theo girlfriend làm chi đâu? Mỗi tối, me cùng pray to God, để cho Ba me dùng có giống Ba you. Néo mà Ba me mà giống Ba you, thì me sẽ mất Ba me. Chừng nào you qua bên A-lá s kà, you nhớ gol cho me nhe. Oh, có cái tâm hình này để me tăng cho you. Tâm
hình này nè, gia đình của me với you đi trên chiếc tàu nè.
(O…Oh… But your Dad is quite a liar. If he loved you and your Mom, too, he would not follow his girlfriend, would he? Every night, I pray to God so that my Dad won’t be the same as yours. If he is, I would lose him. So long as you come to Alaska, don’t forget to give me a ring. Oh, I’ve got you this picture. This picture, you see, you and my family got on board a boat.
- OK. Chúng nào đến đó, me gọi cho you liền.
Er. Me kiểm hình đẹp đẹp hóa, me gọi cho you.
(Okay. As soon as I’ve arrived there, I’ll give you a call instantly.
- Thôi, mình đi zô chơi game đi, đừng có buồn nữa. (So, let’s get in and play games. Don’t be that sad!)

Prepositions, conjunctions are seen rarely transferred. A few interjections, hedges, or discourse markers like guego (well), è ni guê (anyway), ráp bit (rubbish), or ó mai gót (Oh my gosh), só ri (sorry) are predominantly used among the young informants in the study. Most of them are bilingual proficient.

3.3 Integration
From the corpus, most of the lexical items transferred from English share more or less phonological, graphemic, semantic and grammatical integration of the recipient language.

a. Phonological integration: There is a tendency of transferring an English lexeme into Vietnamese by giving it the near-similar form of the original pronunciation or integrating it to a higher or lower degree along a continuum.
High integration means eliminating phonemes that do not exist in Vietnamese, and replacing them with ones appropriate to the recipient system. E.g., Check → /ʃek/, Claim → /leim/, dentist → /dentit/, fix → /fit/.
This often happens when there is no Vietnamese equivalent, and the dropping of the final endings are found popular. However, one of the most striking features in phonological integration in Australian Vietnamese is its falling tonalization of the transferred two lexemes. E.g., tit-sù (tissue), i-zì (easy), ken-sò (cancel), cao-sò (council), K-ipec (Kippax), Giung-ga-lìn (Gungalin). This is true with what Clyne (ibid.) referred as tonemic transference in the case of tone languages such as Chinese Vietnamese and Swedish (Clyne 1985: 104).
b. **Graphemic integration:**

There is a tendency to replace or omit a number of graphemes or clusters of graphemes from the donor language with some existing in the recipient one. This graphemic transference is resulted mostly from phonetic transfer. E.g., Council → written as *kao- só*, Centerlink as *sentò- lin*, manage as *mé nịt*, contract as *ción tràc*, commission as *kóm mít sòng*, settle as *sét tọ*. The tendency of dropping or replacing the final letter with another in the recipient language, however, has been seen by most monolinguals only. Some original forms with partial modification could also be seen as follows:

* Komplit is actually modified, Vietnamized and combined with the original form of “kitchens”

**Semantic integration:**

The data shows a tendency, in nouns mainly, of restructuring the lexical field of the vocabulary integrated.

**Expansion:**

Shop for *shop Tàu* (Asian grocery), *shop thịt* (butcher’s), *shop bánh mi* (bakery), *shop hoa* (flourist’s), *tiền shop* (money paid for shop lease), *mở shop* (open a trading business), *tách/diển form* (detach/fill in the form), *ghếp form* (officially registering for living together)
Compromise form: comprising two morphemes or more of the two languages. E.g., tiền shop, tiền li (lease), tiền ren (rental), tiền côm mít sơn (commission), tiền bon (bond), chéc sô bâng (check the bank account), lánh lum sum (get a lump sum), lấy hố lí dây (take holidays), lấy sít lí (take sick leave), lánh benefít (have family tax benefit), ăn gueo-phe (get welfare benefit), làm gît dôp (doing good job) etc.

The following advert is found on a newspaper about a business sale. Not only semantic compromise (as in sang shop) but also lexicosyntactic transference is found. E.g., trên căn bản walk in walk out (on walk-in-walk-out base).

d. Grammatical integration
Unlike many languages with gender and inflexion (Kaminskas 1972: 86; and Kouzimin 1972: 90 & Clyne 1985: 97), integration in Vietnamese requires no gender, or case and tense forms or assignments of plural forms. That is, transferred nouns and verbs have zero grammatical inflexions.

E.g.,

Hôm qua anh book bác sĩ chưa? (past form)  
Một meat pie và hai sausage roll (plural form)  
Tôi take care (zero particle) việc đó rồi.

This could be accounted for in the language typology of Vietnamese. However, evidence from a number of observations shows that for Vietnamese Australians, the situation is somewhat different. There are some cases with plural and tense form transference into Vietnamese,
Preliminary observations on a “migrant Vietnamese” especially for those whose Vietnamese competency is not regarded as fluent.

3.4 Syntactic transference
Both Rayfield (1970) and Clyne (1976, 1980, 1985) prove that lexical transference tends to proceed from L2 to L1 while phonological transference in the first generation has a reversed direction. As for syntactic transference, the data of the study just shows the direction from L2 to L2, whereas, according to Rayfield and Clyne, it occurs in both directions.

Common forms of syntactic transference are:

a. The use of passive voice in written language and in the spoken language of educated bilinguals:

Eg. - Bà đã đi ra khỏi bệnh viện Manly trong năm ngoái và đi đến Queensland, cuối cùng được tìm thấy bởi Thứ dân Úc tại Coen, thuộc miền Bắc Queensland, ngày 31 tháng 3 năm ngoái.
(She left Manly hospital last year for Queensland and was eventually seen by the Australian Aboriginals in Coen, in the North of Queensland, on last March 31.)

- Người ta tin rằng nó chẳng bao giờ được kiểm tra bởi các nhân viên di trú.
(It is believed that he has never been inspected by immigration officers)
(From Một Phụ Nữ Sydney Bị Giữ Nhâm Lận Trong Trại Di Trú, in Nam Úc Thời Báo, 18/2/05)

- Một quyết định chính thức của Đảng Tự Do sẽ không được đưa ra cho tôi khi cuộc họp của Đảng ở NSW sẽ vẫn để ngày (totally loaned structure)
(An official decision will not be made by The Liberty Party
(Tự Do Không Tranh Cử Werria, Nam Úc Thời Báo, 19/2/05)

b. The use of objects in defining clauses:
- Theo nguồn tin cáo cấp trong đảng Tự Do, đảng không muốn chỉ tiêu cho một cuộc vẫn động mà nó không nghĩ có thể chiến thắng. (in Vietnamese)
(Tự Do Không Tranh Cử Werriwa, Nam Úc Thời Báo, 18/2/05)

c. The English syntactic transference in word order.
- Tôi biết là thư ngắn cảm ơn bà Trummer, ít hàng về việc học
(Short story: Không biết mất, Vietnamese Community Newspaper, 15/7/03)

105
3.5. **Pragmatic transference**

a. **Adoption of patterns of addressing** of colleagues or workmates and acquaintances by first names.

E.g. Tai Van Dang instead of Đặng Văn Tài

b. **The use of pronouns “you” and “me”** as an indication of showing not only solidarity, cooperation but also equality, compassion and sharing (Brown & Gilman 1960, Bate & Benigni 1975, Brown & Levinson 1978, Goffman 1981, Sifianou 1992) especially among 1.5 and second generation children (as shown in 3.2.b). Sometimes, this usage is resulted in a code-switching process which signals a negation of an existing relation or a temporary social relation (Myer-Scotton 1993, Clyne 1994, Ho-Dac 1998).

- Anh T. You (instead of Anh) nói là chiều nay You ở nhà sơn gara, mà sao me (instead of em) không thấy you làm gì cả? (spoken by a wife to husband)

(T., You said that this afternoon you would stay home and paint the garage, but why didn’t I see you did at all?)

(Conversation recorded)

c. **The use of “thank you”** for every little thing offered and the higher frequency of **“sorry”** compared to the traditional vocal habits are also recorded in the corpus. This is salient feature is not only reflected in daily conversation of speech etiquette, but also in every conversational interactions of both formal and informal contexts.

3.6. **Code-switching and syntactic transference**

Similar to Clyne’s findings, a common tendency of switching from Vietnamese to English and vice versa was found among young interlocutors interacting with each other and in the presence of older people respectively. Though this tendency needs further in-depth investigation, the results gained from the study showed that such factors as domain, topic, venue, interpersonal relationship, channel of communication and nature of interaction are conductive to code switching. The most common domains are home and business-related ones where the code switching is strongly marked from Vietnamese to English, whereas the change of topics to home is found in the other way around, whatever venues might be. For some limits, the study did not show clearly whether venue is a striking feature of code-switching. However, what is most noteworthy in the study is that language convergence from Vietnamese to English occurs naturally among young interlocutors, especially when the interpersonal relationship is relatively distant. The more distant it is, the higher the tendency to use English.

In the following “agony” column of a newspaper (Chiếu Dương, issue 24/4/05, p. 32), code switching and syntactic transference from Vietnamese into English several times within both sentences and the whole text, being a linguistic behaviour which is more and more popular in the Viet diasporas.
While friendly transactions are likely to be in Vietnamese, business-related or formal communications are most found in English with high frequency. It is also assumed that the relationship between a home language and English as a host language has emerged as a key determinant of ethnic self-identity (Reitz 1980, Gitelman 1995, Shuval 1998, Remennick 2004) and this is the case with the young educated informants who have rich social capital and are using Vietnamese at home but purely English in their occupational realm. Code switching of this type is often caused by trigger-words (Clyne 1985: 106-107) in the forms of proper nouns, lexical transfers, loan words, and hedges (discourse markers) the most common of which are as well, you know, by the way, you see, alright, okay, anyway, tell you what, you know what etc. and the use of loan pronouns of you and me. In my study, more than 80% of obviously anticipational switching occurs with the “habitual” use of hedges, and loan pronouns and interestingly, the change of discourse topic on to business-related ones. Conversely, the convergence back to Vietnamese language happens only with the presence of non-bilingual elders, especially when the hierarchy must be defined through the use of Vietnamese forms of addressing which are traditionally age-based and family-oriented. In my view, this kind of linguistic and cultural integration process appears to be incremental and rather than replacement-oriented and bears the adhesive or segmented nature of assimilation. Having said so, I would state from the study that there is a tendency for linguistic integration as a response to the demands of the new society, where Vietnamese Australian (of the first and 1.5 generations) would make themselves more functional in their social environment. They would just add new layers to their traditional ethnic identity and lifestyle, but at the same time, try to maintain their cultural values. However, this is not the case with the second generation bilinguals, who develop their
own pathways between home and host lifestyles (Horowitz 2001, Remennick ibid.). This also calls for other in-depth investigations based on a hypothesis that linguistic adaptation and acculturation is contingent with younger age of immigration, higher social capital, social networks and transnational links as shown in our study.

4. Conclusion
This data analysis has shown not only how linguistic integration and transference from Vietnamese as a home language into English as host language occurs, but also how the cultural identity of the interlocutors as first and second immigrants is split: ethnic identity (Vietnamese-Australian) and national identity (Vietnamese). Although the division between the use of English in the public realm and the mother tongue in the private could be stated, English gradually invades informal communications and certainly prevails in business-related domains. Likewise, the Vietnamese spoken by migrants differs from community to community due to socio-cultural elements. The results of the study also shed light on the fact that the acculturation of the Vietnamese immigrants in Australia develops along additive rather than replacement lines, i.e., the Viet immigrants add English (on various levels and in different ways) to their core linguistic and cultural menu, which remains Vietnamese. This “migrant Vietnamese”, in its turn, contains different ethno-linguistic identities, the variations of which are determined by the immigrants’ Vietnamese fluency and competencies in other languages, age of migration, previous contact and nature of contact with the language of the countries of residence. Furthermore, what can be seen through this is that the Vietnamese migrants’ work environment in Australia holds decisive importance in shaping their idiolects and sociolects of Vietnamese, which makes the language spoken in various diaspora groups “a migrant Vietnamese” bearing untold colloquial varieties greatly distant from the standard varieties spoken in the home country.

References
Preliminary observations on a “migrant Vietnamese”


1. Introduction
ALLATIVE case expresses motion to or toward the referent of the noun phrase it marks, which is a ‘schematic vector’ of the path of motion (cf. Talmy 2000: 53-57). The concept ALLATIVE roughly refers to a dynamic spatial relation where a moving entity moves to a reference point. This concept of ALLATIVE entails the whole kinetic PATH schema consisting of SOURCE, PATH and GOAL (cf. Lakoff & Johnson 1999: 32-34). A language may have a number of ALLATIVE markers with different nuances. For example, the English language has several ALLATIVE prepositions: toward (GOAL noun phrase), to (GOAL noun phrase), onto (GOAL noun phrase), and into (GOAL noun phrase).

(1) He walked toward the shop. (PATH-oriented, unbounded path)
(2) He walked to the shop. (GOAL-oriented, bounded path)
(3) He put the files onto a CD. (GOAL-specific, affected target entity)
(4) He put the files into a box. (GOAL-specific, enclosed goal space)

The preposition toward in (1) represents directional or PATH-oriented ALLATIVE implying an unbounded path. The preposition to in (2) represents GOAL-oriented ALLATIVE implying a bounded path. The preposition onto in (3) represents GOAL-specific ALLATIVE implying an affected target entity. And the preposition into in (4) represents GOAL-specific ALLATIVE or ILLATIVE implying an enclosed goal space. Likewise, the Japanese language has two main ALLATIVE particles: (GOAL noun phrase)-e and (GOAL noun phrase)-ni.

(5) kare wa arui-te mise-e mukat-ta.
PRONOUN TOPIC walk-TE shop-toward go-PAST
‘He went toward the shop, walking.’
(PATH-oriented)

(6) kare wa arui-te mise-ni mukat-ta.
PRONOUN TOPIC walk-TE shop-to go-PAST
‘He went to the shop, walking.’
(GOAL-oriented)

The particle -e in (5) represents directional or PATH-oriented ALLATIVE and the particle -ni in (6) represents GOAL-oriented ALLATIVE.

By contrast, the Thai language, I claim, has only one ALLATIVE preposition proper, namely ยัง. This preposition is derived from a verb meaning ‘maintain,’ ‘sustain,’ ‘exist,’ or ‘bring about.’ In modern Thai the lexical item ยัง expresses two grammatical concepts, namely CONTINUOUS aspect and ALLATIVE case. When it precedes a verb phrase, as in (7), it represents CONTINUOUS aspect, that is, to

continue doing or being. When it precedes a noun phrase, as in (8), it represents ALLATIVE case, that is, to or toward a goal.

(7) khāw yan dān rāop ráan
PRONOUN CONTINUOUS walk around shop
‘He was walking around the shop.’
(CONTINUOUS aspect)

(8) khāw dān pay yan ráan
PRONOUN walk go ALLATIVE shop
‘He walked to the shop.’
(ALLATIVE case)

This study focuses on the latter grammatical meaning of yan, namely ALLATIVE. The aim of this study is to attest to yan’s grammatical status as ALLATIVE preposition by examining actual tokens gathered from corpus data including published literary works, magazines and newspapers.

This paper is organized in the following way. In Section 2, I will make some remarks on the nature of yan to which little attention has hitherto been given. In Section 3, I will point out some distinctive features lying between the ALLATIVE preposition proper yan and ARRIVAL verbs in order to show their different nature and use. And Section 4 provides concluding remarks.

2. The nature of ALLATIVE preposition yan
Let us start with outlining the nature of the Thai ALLATIVE preposition yan.

2.1. Non-obligatory ALLATIVE marker
yan is not an obligatory ALLATIVE marker. It is observed that motion expressions with the ALLATIVE preposition yan abound in literary works, but the expressions are seldom used in everyday conversation. This is partly because without the overt ALLATIVE marker, a default ALLATIVE sense of Thai motion expressions could emerge from serialization of a PATH verb (e.g. pay ‘go’) and a GOAL noun or prepositional phrase, as illustrated in (9) and (10).

(9) khāw dān kun pay [yǒt khāw]
PRONOUN walk ascend go mountaintop
[GOAL Noun Phrase]
‘He walked up to the mountaintop.’

(10) khāw dān kun pay [bōn yǒt khāw]
PRONOUN walk ascend go on mountaintop
[GOAL Prepositional Phrase]
‘He walked up to the mountaintop.’
2.2. Specific ALLATIVE marker
The ALLATIVE preposition  yang  indicates the endpoint of an ‘axial motion’ which is forward motion of an entity along an axial path, as exemplified in (8) above. The prepositional phrase beginning with  thuŋ  can indicate the endpoint of an axial motion as well, as in (11).

(11) khâw  døn  pay  cãak  bân  thuŋ  røoŋ  rian
PRONOUN walk  go  from house  to  school
‘He walked from his house to the school.’
(Axial motion)

The endpoint of a ‘horizontal or vertical motion’ which is the shift of the head of a line extending from a fixed point along a horizontal or vertical path, on the other hand, is marked by  hâa, as in (12), and  carôt,  cõt  or  thuŋ, as in (13).

(12) khâw  hân  nâa  khâw  hâa  dêk
PRONOUN turn  face  enter  to  child
‘He turned his face to the child.’
(Horizontal motion)

(13) mñoŋ  khâw  cãak/ tân  têc  hûa  carôt/ cõt/ thuŋ  thâw
look  PRONOUN from head  to  foot
‘(She) looked at him from head to toe.’
(Vertical motion)

At any rate, these lexical items (or ARRIVAL verbs) are not prepositions ‘proper’ but verbs that are capable of serving as prepositions in certain contexts. We shall return to this point in Section 3.4.

2.3. Fully grammaticalized ALLATIVE marker
The grammatical concept ALLATIVE expressed by  yang  is distinct from the verbal concept ARRIVAL expressed by a variety of ARRIVAL verbs (e.g.  thuŋ  ‘arrive,’  khâw  ‘enter,’  sây  ‘put in,’  pathâ  ‘collide,’  yût  ‘halt,’  thâap  ‘lay flat against,’  sûu  ‘arrive and share/stay,’ etc.). Consider a contrastive pair of examples in (14) and (15).

(14) a. khâw  døn  khûn  pay  thuŋ  yôt  khâw
PRONOUN walk  ascend  go  arrive  mountaintop
‘He walked up and arrived at the mountaintop.’
(Complex event)

b. khâw  døn  khûn  pay  (yang)  yôt  khâw
PRONOUN walk  ascend  go  (to)  mountaintop
‘He walked up to the mountaintop.’
(Simplex event)
(15) a. maa yût thîi ráañ
come halt at shop
‘(He) came and stopped at the shop.’
(Complex event)

b. maa (yañ) thîi ráañ
come (to) at shop
‘(He) came to the location of the shop.’
(Simplex event)

(14a) and (15a) include an ARRIVAL verb (thuûng ‘arrive,’ yût ‘halt’) while (14b) and (15b) do not. The former expressions have a complex event structure consisting of a motion and an arrival as a result, which I call ‘ARRIVAL event’ (cf. Takahashi, to appear), while the latter expressions have a simplex event structure consisting of a single motion to an endpoint. In this study the concept ARRIVAL is defined as an event of a moving entity’s reaching an endpoint as a result of its prior motion along a path toward the endpoint. The concept ARRIVAL may involve the characteristic of a goal entity and/or the type of effect resulting from the arrival. Whereas the concept ALLATIVE is highly schematic and purely directional, the concept ARRIVAL is fairly rich in the content of event. In other words, the former is grammatical and semantically bleached; the latter is lexical and contentful.

3. ALLATIVE preposition yañ vs. ARRIVAL verb sùu
To clarify the grammatical status of yañ as ALLATIVE preposition proper, in this study I will compare yañ with sùu as a representative of ARRIVAL verbs. The meaning of sùu is somewhat complex, namely ‘arrive and share/stay,’ as exemplified in (16).

(16) khâw kêp mańkhût pay sùu phûán bân
PRONOUN gather mangosteen go arrive and share/stay neighbours
‘He harvested mangosteen and went to share them with his neighbors.’ (Tomita 1990: 1835)

In (17) sùu follows a series of locomotion verbs. In such a context, sùu is often considered as an ALLATIVE preposition.

(17) lança niñi khâw pay sùu ?aaraam phrà?phûtthacâw
run flee enter go arrive and share/stay temple
‘(He) ran away and got to the temple and stayed in.’
> ‘(He) ran away into the temple.’

Indeed, sùu can function as a preposition. However, as will be discussed later, it is the case with some marked contexts such as when being in combination with another particular preposition (e.g. câak kruûngthêep sùu manîlâa ‘from Bangkok to Manila’) or when used in more abstract domains than the spatial domain (e.g. sùu ?aanaakhôt ‘into the future’).

Based on my examination of corpus data including many tokens of yañ and sùu, I argue that yañ has undergone the process of grammaticalization much further than sùu.
Noss (1964: 148), Kölver (1984: 16), Diller (2001: 165) among others regard both yaŋ and sūu as fully grammaticalized ALLATIVE prepositions. But I do not share this common view. I would claim that yaŋ has become an ALLATIVE preposition and lost its content meaning but sūu has not yet done so. sūu largely functions as an ARRIVAL verb. The following observations serve as a piece of evidence in favor of my claim.

3.1. Non-preposition-like behavior of sūu
First, while yaŋ is always followed by a noun phrase, sūu can take place at the final position of serial verb constructions, as in (18).

(18) năn khuu ṭhīi rāap lē? yōot kḥaw
that COPULA place flat and hilltop
sūu phūak khaw praathanāa cā?
RELATIVE PRONOUN people PRONOUN wish MODAL
hōon tua sūu
sway while hanging body arrive and share/stay
‘That is a flat place and a hilltop, where they wish to lift themselves and get to and stay.’

In this regard, other ARRIVAL verbs behave in the same way. (19) to (21) give other examples ending up with an ARRIVAL verb, namely sāy ‘put in,’ hāa ‘seek’ and thūŋ ‘arrive.’

(19) yaa hāy khīi fūnlōn lōŋ sāy
don’t CAUSATIVE dust fall descend put in
‘Don’t let dust fall down and go into (something).’

(20) man lēn pay hāa
PRONOUN run go seek
‘It ran away and approached (something).’

(21) raw caŋ? bēek sāŋkhāan sāpparaŋkhee nīi
PRONOUN MODAL carry body decrepit this
dānthāaŋ pay thūŋ
travel go arrive
‘We will travel with these decrepit bodies of us and reach (some place).’

3.2. Linear order constraint on the combination of sūu and yaŋ
Secondly, yaŋ and sūu may co-occur. When they co-occur, as in (22), sūu as an ARRIVAL verb must precede yaŋ as an ALLATIVE preposition. The same observation applies to (23) where thūŋ, an ARRIVAL verb, precedes yaŋ.

(22) kḥaw pay sūu yaŋ ṭhīi kļāy
enter go arrive and share/stay to place near
‘(The sunlight) went inside, reached and stayed, (to) a near place.’
Kiyoko Takahashi

(23) khabuan thahaan daen suan sanam maa thun yaj
marching line soldier move pass field come arrive to
?anusawarii monument
‘The soldiers marched across the field and reached the monument.’ (Intratat 1996: 128)

This is due to syntactic principles of Thai grammar. Considering the fixed linear order of the constituents of a verb phrase, namely verbs are followed by a prepositional phrase, we can say that suu and thun in the preceding position are more verb-like and yaj in the following position is more preposition-like.

3.3. ARRIVAL event denoted by suu
Thirdly, the usage of the ARRIVAL verb suu always suggests a ‘stative’ and ‘resultative’ situation after a given arrival, as exemplified in (24) to (26).

(24) khaw klap suu baan
PRONOUN turn back arrive and share/stay house
lan daen
CLASSIFIER originally
‘He returned and got to the original house and stayed.’

(25) phaa com lon suu kon samut
lead sink descend arrive and share/stay bottom sea
‘(It) led (something) down and got to the bottom of the sea and stayed.’

(26) loay khoom kun suu thong faa
float lantern ascend arrive and share/stay sky
‘(They) sent up floating-lanterns which got to the sky and stayed.’

This implication of ‘stasis’ is essentially connected with the verb’s original meaning, that is, to get to some place and share something with someone, as mentioned before.

From the tokens of suu I gathered, the following scenario of an ARRIVAL event can be generalized: After arrival, the moving entity in question settles in the arrival space; otherwise, the moving entity virtually fades into some space as it goes away from the observer or the moving entity is fused into the setting to the extent that its motion cannot be detected. This is the prototypical ARRIVAL event that suu denotes.

3.4. Using suu as ALLATIVE preposition in a limiting case
Actually, suu can function as a preposition only when the prepositional phrase led by suu is in combination with an ABLATIVE prepositional phrase (e.g. caak … suu … ‘from … to …’), as in (27), or in adjacency to the ARRIVAL verb khaw ‘enter’ (i.e. khaw suu … ‘move into …’), as in (28). 

116
The allative preposition in Thai

(27) deaux thaan caak krungtheep suu manila
travel from Bangkok to Manila
‘(They) traveled from Bangkok to Manila.’

(28) sdee sathao klap maa khaw suu taa khoong
light reflect turn back come enter to eye of
phu mooi viewer
‘The reflected light came into the eyes of the viewer.’

The serialization of the ARRIVAL verb khaw and the preposition suu brings about an elaborated sense of the ALLATIVE, namely the concept of ILLATIVE or ‘motion into some place.’ Especially, the serialization of khaw and suu denotes a special type of ILLATIVE that involves the notion of resultant stasis.

Additionally, uses of other ARRIVAL verbs as an ALLATIVE preposition are also subject to the collocational circumstances. That is, the prepositional phrase beginning with haa must follow the arrival verb khaw ‘enter,’ as in (29)=(12), and the prepositional phrase beginning with carot, cot or thuung must be combined with an ABLATIVE prepositional phrase beginning with caak or tanj tce ‘from,’ as in (30)=(13) and (31)=(11).

(29) khaw haa naa khaw haa dek
PRONOUN turn face enter to child
‘He turned his face to the child.’

(30) mooi khaw caak/ tanj tce hua carot/ cot/ thuung thaw
look PRONOUN from head to foot
‘(She) looked at him from head to toe.’

(31) khaw dään pay caak banh thuung roon rian
PRONOUN walk go from house to school
‘He walked from his house to the school.’

In more abstract domains than the spatial domain, such as in the temporal domain, suu as well as other ARRIVAL verbs can be used as a preposition, as in (32) and (33) below, but further investigation into usages of ARRIVAL verbs in such abstract domains is beyond the scope of this study.

(32) song toê caak run nueng suu ?iik
send connect from generation one to another
run nueng generation one
‘(They) sent (something) from one generation to another.’

(33) mooi fi kaaaw nay sêen thaan suu ?aanaakhôt
look the action of taking a step in path to future
‘(He) looked at making steps along the path to the future.’
To summarize this section, my proposal is that whereas ยน จ has become an ALLATIVE preposition, สู ข ‘arrive and share/stay’ and other ARRIVAL verbs largely function as a verb designating an ARRIVAL event. Specifically, ยน highlights an axial path toward an endpoint and สู expresses an ARRIVAL event involving resultant stasis.

4. Conclusion

The Thai language has a number of lexical items used as both a content word (e.g. verb) and a function word (e.g. preposition). ARRIVAL verbs like ทูน ‘arrive’ and สู ‘arrive and share/stay’ are also used as ALLATIVE markers. This study has shown, however, that in fact these ARRIVAL verbs mainly express quite a substantial event of ARRIVAL and they are allowed to function as an ALLATIVE marker in a limiting case. On the contrary, the lexical item ยน no longer expresses a verbal (or contentful) meaning in present-day Thai, but exclusively represents a grammatical (or non-contentful) meaning, more precisely, a temporally or spatially schematic meaning, i.e., either CONTINUOUS aspect or ALLATIVE case. It should be concluded, from what has been discussed in this paper, that there is only one ALLATIVE marker proper in Thai, that is, ยน.

In passing, Thai speakers employ ‘Accomplishment Constructions’ consisting of CAUSE and EFFECT components to delineate temporally bounded events with the latter EFFECT part highlighted (cf. Takahashi, in press), as illustrated in (34) to (36) below. In (34) someone’s action of lifting a bag (CAUSE) leads to its upward movement (EFFECT). In (35) someone’s stretching his line of sight away (CAUSE) brings about his visual perception of mountains (EFFECT). In (36) something’s falling off (CAUSE) results in its broken state (EFFECT). The speakers uttering these sentences should regard the posterior EFFECT sub-event as a natural consequence of the prior CAUSE sub-event and concern themselves with the realization of the EFFECT sub-event.

(34) [ยอ ก แปร] [ขุน] 
lift bag ascend 
[CAUSE] [EFFECT] 
‘(He) lifted a bag and it moved upward.’

(35) [มณ ปาย] [เห็น ภูเขา] 
look go see mountain 
[CAUSE] [EFFECT] 
‘(He) looked away and caught sight of mountains.’

(36) [ตอ] [เทเกก] 
go down be broken 
[CAUSE] [EFFECT] 
‘(It) fell off and was broken.’

My basic view regarding ARRIVAL expressions, as in (37) and (38) below (as well as (14a), (15a), (16) to (26) and (28) above), is that they are a kind of Accomplishment Construction and that an ARRIVAL verb in the following EFFECT
The allative preposition in Thai

cOMPONENT of the construction (ثوون ‘arrive’ in (37) and سوو ‘arrive and share/stay’ in (38)) denotes a substantial event of ARRIVAL that eventually arises from a prior locomotion event designated by PATH verb(s) in the preceding CAUSE component.

(37) [婵 เน] [ثوون ราแนน]
    walk go arrive shop
    [CAUSE] [EFFECT]
    ‘(He) walked away and reached the shop.’

(38) [ขุน เน] [สูู ชัน บอน]
    ascend go arrive and share/stay upstairs
    [CAUSE] [EFFECT]
    ‘(He) went up and got to the upstairs and stayed.’

The speaker of (37), for instance, conceptualizes that the unexpressed mover’s arrival at the shop occurs as a natural result of his walking away previous to the arrival. Crucially, the speaker’s focus of attention is placed on the realization of the posterior EFFECT sub-event (his resultant arrival at the shop) rather than the prior CAUSE sub-event (his relocation toward the shop) which is a precondition for the realization.

In contrast, expressions for simplex locomotion event, as in (39) and (40) below (as well as (8) to (11), (14b), (15b), (27) and (31) above) which may or may not contain the ALLATIVE marker ยัน, make no clause-internal separation of CAUSE and EFFECT components.

(39) 婵 เน (ยัน) ราแนน
    walk go (to) shop
    ‘(He) walked to the shop.’

(40) ขุน เน (ยัน) ชัน บอน
    ascend go (to) upstairs
    ‘(He) went up to the upstairs.’

Although the physical state of affairs described by (37) and (39) may be identical, the speaker’s conceptualization is different from each other. Unlike the speaker of (37), the speaker of (39) views the motion event in question as a simplex one: the unexpressed mover relocates on foot in the direction to or toward the shop and concurrently away from a certain reference point.

Note
I would like to thank Robert De Silva for helpful comments on the early version of this paper and stylistic suggestions. The remaining faults are all my own.

1. The data used for this study were collected mainly from a number of published literary works I randomly selected and partly from a computerized corpus of the Thai language that belong to the National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Ministry of Science, Technology and Environment, Thailand.
References


Kölver, Ulrike. 1984. Local Prepositions and Serial Verb Constructions in Thai, in Arbeiten des kölner universalien project, Nr.56.


PRONOUN RETENTION IN KHMER AND THAI RELATIVE CLAUSES

Natchanan Yaowapat
Department of Linguistics
Chulalongkorn University
Natchanan.y@student.chula.ac.th

Abstract
The construction to be investigated here is the ‘Relative Clause’ in Khmer (Cambodian) and Thai. This study aims to consider only one aspect or one characteristic of relative clauses in both languages, namely, pronoun retention. The occurrence of resumptive pronouns, pronouns which are coreferential with head noun phrases, in relative clauses will be investigated and some of the constraints on their occurrence in the two languages will be discussed.

1. Pronoun Retention
According to Comrie (1981), pronoun retention is typologically a way to encode the role of the head noun in the embedded sentence (relative clause). The head noun remains explicitly in the embedded sentence in pronominal form. Pronoun retention is one strategy used to form relative clauses in many languages. An example of pronoun retention in non-standard English is *this is the road that I know where it leads*. In this example, the relative clause is *that I know where it leads* functions to modify the head noun phrase *the road* in the main clause. The resumptive pronoun *it* in the relative clause refers to the head noun *road* and it remains in the normal position of the clause subject, the grammatical relation that it encodes.

The following examples illustrate relative clauses formed by pronoun retention in Khmer and Thai.

Khmer
(1) *kjom skoal kruu m-neak [dael koat bagrien phiesaa phaklei]*
   ‘I know teacher one-person REL s/he teach language English’

   The noun phrase being relativized in the above example is *kruu mneak* ‘a teacher’. The resumptive pronoun *koat* ‘s/he’ which is coreferential with the noun phrase *kruu mneak* occurs in the normal position of the subject in the relative clause, that is, preceding the main verb *bagrien* ‘teach’(the basic word order in Khmer is S-V-O).

1 I would like to express my deep appreciation to the Thailand Research Fund for sponsoring me to develop this paper under the supervision of Professor Bernard Comrie at the Department of Linguistics, University of California Santa Barbara. I am especially grateful to Professor Bernard Comrie and Professor Amara Prasithrathsint for their supervision and advice.

The pronoun khāw ‘s/he’ in the relative clause above is coreferential with the head noun khruu ‘teacher’ in the main clause. The grammatical role of that pronoun is subject of the embedded clause and the pronoun occurs in the subject position, that is, precedes the verb sǒon ‘teach’ (the basic word order in Thai is also S-V-O).

With the pronoun retention strategy, it is found that it allows more NPs to be relativized. Some languages like Basque don’t normally allow relativization on Genitive NPs. But with this strategy, Genitive NPs are relativizable. (Keenan 1985).

2. Relative Clauses in Khmer and Thai

2.1 Relative Clauses in Khmer

Relative clauses in Khmer are marked by the word dael, the general marker for linking subordinate clauses with head nouns (Comrie & Horie 1995). Typologically, relative clauses in this language are of the postnominal type, that is, relative clauses occur following head NPs, as in (3).

(3) kʰaek tʰewahslaaphæ sʰamæv tii moatstwaŋ muŋy [dael riŋtʰukʔah]
crow fly toward river one REL dry
‘The crow flies to a river which is dry.’ [written text]

In (3), the relative clause which modifies the head NP tii moatstwaŋ muŋy ‘a river’ is dael riŋtʰukʔah ‘which is dry’. The marker dael occurs clause-initially and the relative clause follows the head NP.

Besides the pronoun retention strategy, relative clauses in Khmer can be formed by the gap strategy, the least explicit way of encoding the role of the head NP in embedded clauses. Instead of having a personal pronoun remaining in the relative clause, there is a missing NP which is coreferential with the head noun. In example (3) above, from the basic word order S-V-O and from the argument structure of the main verb in the embedded clause, there is a missing argument in the relative clause dael riŋtʰukʔah ‘which is dry’, that is, the subject NP of the clause. Since the missing noun phrase in the relative clause formed by the gap strategy is coreferential with the head NP, the missing subject argument in the example (3) can be retrievable as tii moatstwaŋ muŋy ‘a river’.

2 Although relative clauses in Khmer and Thai allow the omission of the relative clause markers, the present study includes only relative clauses with the overt markers.

Both written and spoken data were collected in the two languages. Examples from written texts are marked as [written text], all other examples are elicited spoken examples. There are around 150 relative clauses collected in each language.

Restrictive and non-restrictive types are not distinguished in the study.

3 The written data were collected from newspapers, journals, and short stories. The spoken data were elicited from two native Khmer informants in Thailand, one male graduate student and one female graduate student at Chulalongkorn University)
Pronoun retention in Khmer and Thai relative clauses

Concerning the NP positions that can be relativized in Khmer, subjects, direct objects, indirect objects and possessors can all be relativized. Subjects and indirect objects can be relativized by using either the gap strategy or pronoun retention. Direct objects are relativized only by the gap strategy whereas possessive NPs can be relativized only by pronoun retention. The examples (4) – (7) will illustrate the relativization of those NP positions.

(4) သមល ឈ្មោះពាក្យ កាត់ថ្មី បារីរោង ឆ្លាញ [ៗ នេះ ឆ្លាញ មិន រឹត] ទំលៃថ្មី
day one fox invite cat REL be friend visit
ស្រគ១ កាល រឹត កេឡុង
village birth belong to him
‘One day, the fox invites a cat which is his friend to visit his hometown.’
[written text]

The subject NP ឆ្លាញ ‘cat’ is relativized and leaves a gap in the subject position of the relative clause.

(5) ក់ន៊ីក់ពី១ នីក្នុងមិន ជាមួយ ២ អេក្រង់ ឆ្លាញ មិន គួរ ធ្វើ
daughter their two CLF not know do exercise REL person
គុណ ដែល អេក្រង់ មិន នឹង ពីរាំ
teacher assign work at house
‘Their two daughters don’t know how to do the exercise which the teacher assigns to be done at home.’
[written text]

The direct object NP ឆ្លាញ ‘exercise’ is relativized in the above example and leaves a gap in the direct object position, that is, after the verb មិន នឹង ពីរាំ ‘work’.

(6) ក្នុង [១ឃាំ ក្រុម នីក្នុង ៥ ឆ្នាំ មានចំនួន] ឈ្មោះ នេះ ពីរាំ នឹង សម្រាប់
child REL I give money age five year
‘The child to whom I gave some money is five years old.’

The indirect object ក្នុង ‘child’ is relativized and leaves a gap in the clause-final position of the relative clause ១ឃាំ ក្រុម នីក្នុង ៥ ឆ្នាំ មាន

(7) កាត់ [១ឃាំ មិន ក់ពី១ ក់ពី១ ឆ្លាញ មិន នឹង ពីរាំ]
chair REL leg it PAST break be chair of him
‘The chair the leg of which leg is broken is his chair.’

The possessor កាត់ ‘chair’ is relativized. The personal pronoun វី ‘it’ occurs in the relative clause to encode the possessive role of the head noun កាត់ ‘chair’.
2.2 Relative Clauses in Thai

Unlike relative clauses in Khmer, relative clauses in Thai can be introduced by one of the three markers, namely, \( \text{thi} \), \( \text{s} \), and \( \text{?an} \). The \( \text{thi} \) marker is the most frequent and neutral choice whereas \( \text{s} \) is used in more literary and formal style. \( \text{thi} \) and \( \text{s} \) are mostly used interchangeably. The \( \text{?an} \) marker is quite archaic. It is still used in present day Thai as the least frequent choice.

Relative clauses in Thai are also typologically of the postnominal type and can be formed by the two strategies, the gap strategy and the pronoun retention strategy. Unlike Khmer, subjects, objects (direct and indirect objects), and possessors can be relativized by either the gap strategy or pronoun retention. However, the gap strategy is more frequently used in both written and spoken Thai. The examples (8) – (11) will illustrate the relativization of all possible NPs.

(8) \( \text{khun hën thi} \text{idin [ thii } \text{Ø } \text{ît kâp dûan tawan} \text{tok khôŋ raw]} \text{mây yu} \text{see land REL adjoin with side west of us Q} \)

‘Do you see the land which is next to our west side?’ [written text]

The subject NP \( \text{thi} \text{idin} \) ‘land’ is relativized. The gap in the subject position, preceding the verb \( \text{ít} \), in the embedded clause is coreferential with that NP.

(9) \( \text{phuakkhôw yâŋ mây khun phînthî hûrîween sùn lumphinii [ sîŋ chây they still not return area around garden Lumpini REL use pen sàmnàkyaan chûakhraaw }] \)

\( \text{Ø be office temporary} \)

‘They haven’t returned the area around the Lumpini Garden which is used as the temporary office.’ [written text]

The direct object \( \text{phînthî hûrîween sùn lumphinii} \) ‘the area around the Lumpini Garden’ is relativized and then leaves a gap in the direct object position in the relative clause, that is, after the verb \( \text{chày} \) ‘use’.

(10) \( \text{dek [ thii phô̂mêe hây θøn θâayâayy] màk thuîk taamcay child REL parents give money easily often PASS spoil} \)

‘The child to whom the parents easily give money is often spoiled.’

The indirect object \( \text{dek} \) ‘child’ is relativized and then leaves the gap after the direct object \( \text{θøn} \) ‘money’ in the relative clause.

(11) \( \text{chân rûcâk phûuchaay [thii phanrayaa θìay]} \)

\( \text{I know man REL wife sick} \)

‘I know the man whose wife is sick.’

---

\(^4\) The written data were collected from the online Thai corpus (newspapers, magazines, short stories, etc.) via the program Thai Concordance Online by the Department of Linguistics, Chulalongkorn University. URL: http://www.arts.chula.ac.th/~ling/ThaiConc

The spoken Thai data used in this study were elicited in the workshop on relative clauses in Thai at Chulalongkorn University, May 2003.
Pronoun retention in Khmer and Thai relative clauses

The possessor  phûuchaaay ‘man’ is relativized and the gap after the possessed NP phanrayaa ‘wife’ is coreferential with that noun.

3. Pronoun Retention in Khmer
3.1 Pronoun Retention in Simple Sentences

Pronoun retention, likewise the gap strategy, can be said to be the ‘primary relativization strategies’ in Khmer since they can be used to relativize subject NPs, the most easily accessible NPs (Keenan & Comrie 1977). However, pronoun retention with relativized direct object is unacceptable to native speakers. Like the relativized subjects, relativized indirect objects are found with pronoun retention and the occurrence of the resumptive pronoun is optional. On the other hand, the occurrence of a resumptive pronoun is obligatory for relativized possessive NPs. The examples (12) – (15) will illustrate these findings.

(12) (subject)
k้อย skoal m.muh klah [dael kei tɔɔ-kaa nun tii nuh] I know human some REL s/he work at place there ‘I know someone who works there.’

(13) (direct object)
* ckae [ dael kploy srə-lap vie ] baan slap dog REL I love it PAST die ckae [ dael kploy srə-lap Ø ] baan slap (Gap) dog REL I love PAST die ‘The dog which I love a lot has died.’

(14) (indirect object)
kmeeg [ dael kploy ?aoy luy vie ] ?aayu ? pram cnam child REL I give money it age five year ‘The child to whom I gave some money is five years old.’

(15) (possessor)
boʔraʔ [ dael ckae (ʔɔʔəh) koat rɔwbat ] mun sbaay-caʔ man REL dog of him run away not happy * boʔraʔ [ dael ckae (ʔɔʔəh) Ø rɔwbat ] mun sbaay-caʔ man REL dog of run away not happy ‘The man whose dog ran away is sad.’

The distribution of pronoun retention in Khmer is particularly interesting from a more general theoretical perspective. According to Keenan and Comrie (1977), if the primary relativization strategy, in this case ‘pronoun retention’, in a language is possible for a particular position on the Accessibility Hierarchy, then it must also be possible for all positions lower on the Accessibility Hierarchy. In other words, in the pronoun retention case, the higher NPs on the hierarchy imply the lower NPs.
From the Noun Phrase Accessibility Hierarchy proposed by Keenan & Comrie above, Direct Object is lower in position than Subject, Indirect Object is lower than Direct Object and Subject, and so on. The lower NP on the hierarchy is, according to the claim, more accessible, in this case easier to relativize, than the higher NP. So Subject is the most accessible NP, in this case, is the easiest NP to relativize. Direct Object is more accessible than Indirect Object and so on. Oblique, according to the hierarchy, means the major Oblique case NP which expresses an argument of the main predicate, as \textit{the table} in \textit{Janie put the glass on the table}. The Oblique here doesn’t mean the one with the adverbial function as \textit{in China} in \textit{Wen Ling studies in China}. Oblique NPs in Khmer and Thai are mostly in adverbial function so Oblique is excluded in the present study.

With respect to pronoun retention in Khmer, the strategy applies to Subject, so it should apply to all other lower NPs, that is, Direct Object, Indirect Object and Genitive. However, pronoun retention in Khmer is a counterexample to this particular claim since pronoun retention doesn’t apply with a relativized Direct Object, although this is lower than Subject. Yet, it starts to apply again to Indirect Object and Possessor. In addition, pronoun retention in Khmer also violates the first Hierarchy Constraint which is stated that ‘any relative clause forming strategy must apply to a continuous segment of the Accessibility Hierarchy.

According to Keenan (1985), the lower an NP is on the hierarchy, the more common it is to find it expressed by pronouns. That is, Indirect Object is more commonly expressed by a pronoun in a relative clause than Direct Object. Direct Object more commonly encodes the role of the head NP in the relative clause than Subject. Khmer, again, provides a counterexample to this generalization.

Concerning pronoun retention with relativized subjects in Khmer, it is found that not all subjects can be encoded by the resumptive pronouns. Only animate subjects like humans and animals can be relativized with the pronoun retention strategy, as in the examples below. (Note that relativizing a possessor, where pronoun retention is obligatory, does allow, indeed require a resumptive pronoun even with inanimate heads, as in example (7).)
Pronoun retention in Khmer and Thai relative clauses

(16) **boʔrah** [dael koat cuv kmeen] cie taa rboah kpom
    man REL he help child be grandfather of I
    ‘The man who helps the child is my grandfather.’

(17) **ckae** [dael vie kampuy deek] cie ckae rboah kpom
dog REL it PROG sleep be dog of I
    ‘The dog which is sleeping is my dog.’

(18) **mek** [dael bak ]
    branch (of tree) REL break
    ‘The branch which is broken’
    * **mek** [dael vie bak ]
    branch REL it break

(19) **kantray** [dael cak kradaah ]
    scissors REL prick paper
    ‘The scissors which prick the paper’
    * **kantray** [dael vie cak kradaah ]
    scissors REL it prick paper

(20) **laan** [dael cool kmeen ]
    car REL collide child
    ‘The car which hits the child’
    * **laan** [dael vie cool kmeen ]
    car REL it collide child

(21) **pteah** [dael nuv khaet siemreap ]
    house REL locate/situate province Siem Reap
    ‘The house which is situated in Siem Reap’
    * **pteah** [dael vie nuv khaet siemreap ]
    house REL it locate/situate province Siem Reap

From example (16) – (21), the NP boʔrah ‘man’ and ckae ‘dog’ can be relativized with the pronoun retention strategy since they are animate, whereas the NP mek ‘branch of tree’, kantray ‘scissors’, laan ‘car’ and pteah ‘house’ can not be relativized since they are inanimate.

3.2 Pronoun Retention in Complex Sentences
This section will illustrate the occurrence of the resumptive pronoun with relativized NPs in subordinate clauses, such as in verb complement clauses. In Khmer, the occurrence of the coreferential personal pronouns in subordinate clauses follows by and large with that in simple sentences. That is, subjects, indirect objects, and possessors in subordinate clauses can be relativized by pronoun retention whereas with direct objects this is impossible, as in the example below.
(22)

a. ต้นล้าน ยาاوي แพต
ธิต ท้า
กรม หรู ราช รี เค มาย เค เค เค
ที่คุณ
คุณ
ที่คุณ
คุณ
คุณ
คุณ
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริย์
กษัตริ�
Pronoun retention in Khmer and Thai relative clauses

The resumptive pronouns in the subject and indirect object positions are optional with relativization in simple sentences but they are obligatorily present with relativization in subordinate clauses as in 22b. and 23b. The reason may be that the retrieval of the information in the simple sentences is easier than in the complex sentences. The obligatory occurrence of the resumptive pronoun with relativized NPs in complex sentences may help the hearers to retrieve the information.

4. Pronoun Retention in Thai
4.1 Pronoun Retention in Simple sentences

Pronoun retention in Thai, and likewise the gap strategy, are also the ‘primary relativization strategies’ since they can be used to relativize subject NPs. Relative clauses formed by the pronoun retention strategy are mostly found in colloquial speech, and are rarely found in written texts, whereas the relative clauses formed by the gap strategy are found in both colloquial and written styles. Pronoun retention in Thai is much more flexible than in Khmer. The resumptive pronouns can occur if the relativized noun phrase is any of subject, direct object, indirect object, or possessor, as in the examples below.

(24) (subject)
\[ māa [thīi man kin khēnsīam] mák cāʔ khēŋ-reeʔ \]
dog REL it eat calcium often will healthy
‘Dogs which take calcium are often healthy.’

(25) (direct object)
\[ māa [thīi khun rák man máʔk] taay léw \]
dog REL you love it much die already
‘The dog which you love a lot has died.’

(26) (indirect object)
\[ dēk [thīi čhān háy ḏen (kēe) khāw] ʔaayuʔ háa khūap \]
child REL I give money to s/he age five year
‘The child to whom I gave some money is five years old.’

(27) (possessor)
\[ phūuchaay [thīi māa (khōoŋ) khāw nī pay] kamlaŋ saw \]
man REL dog of s/he run go PROG sad
‘The man whose dog ran away is sad.’

Since Thai allows pronoun retention with all types of NPs, Thai doesn’t violate any NP Accessibility Hierarchy claim, constraint, or generalization proposed by Keenan & Comrie.

Unlike Khmer, there is no constraint on relativizing subject NPs with pronoun retention. All kinds of NPs, animate or inanimate, can be encoded in the relative clauses by the personal pronouns, as in the examples below.
Although the occurrence of pronoun retention in Thai doesn’t have any constraint relating to the semantic properties of the relativized NPs, it seems to have some limitation with the occurrence of the resumptive pronoun relating to the choice of the relative markers. It is found that relative clauses introduced by the ‘?an marker, mostly found with relativized inanimate subjects, do not allow the resumptive pronouns to occur, whereas there seems to be no restriction with relative clauses introduced by the ‘thi’ marker and the ‘s?h’ marker, as illustrated in the examples below.

(33) a. * kh?w d?ay h?y neewkh?it [?an man pen prayoot s?amr?p th?uk khon] he PAST give idea REL it be benefit for every people ‘He gave an idea which is beneficial for everyone.’

b. kh?w d?ay h?y neewkh?it [?an ?h pen prayoot s?amr?p th?uk khon] he PAST give idea REL be benefit for every people ‘He gave an idea which is beneficial for everyone.’

c. kh?w d?ay h?y neewkh?it [s?h man pen prayoot s?amr?p th?uk khon] he PAST give idea REL it be benefit for every people ‘He gave an idea which is beneficial for everyone.’

d. kh?w d?ay h?y neewkh?it [thi man pen prayoot s?amr?p th?uk khon] he PAST give idea REL it be benefit for every people ‘He gave an idea which is beneficial for everyone.’
Pronoun retention in Khmer and Thai relative clauses

(34)
a. *niŋ pen raŋwan [ʔan man yīŋŋay thīsūt nay chiwít]
   this be prize/reward REL it big/great superlatively in life
   ‘This is the prize which is the biggest in (my) life.’

b. niŋ pen raŋwan [ʔan Ø yīŋŋay thīsūt nay chiwít]
   this be prize/reward REL big/great superlatively in life
   ‘This is the prize which is the biggest in (my) life.’

c. niŋ pen raŋwan [thī man yīŋŋay thīsūt nay chiwít]
   this be prize/reward REL it big/great superlatively in life
   ‘This is the prize which is the biggest in (my) life.’

d. niŋ pen raŋwan [sīŋ man yīŋŋay thīsūt nay chiwít]
   this be prize/reward REL it big/great superlatively in life
   ‘This is the prize which is the biggest in (my) life.’

So far, there seems to be no limitation for pronoun retention with relative clauses introduced by thīi and sīŋ no matter whether the relativized NPs are subjects, objects, or possessors, and no matter whether the relativized NPs are animate or inanimate, concrete or abstract.

4.2 Pronoun Retention in Subordinate Clauses
The resumptive pronouns can encode the role of relativized subjects, direct objects, indirect objects and possessive NPs in subordinate clauses. Unlike Khmer, the occurrence of the resumptive pronouns is optional for relativized subjects, direct objects, indirect objects. With respect to possessors, some relative clauses with relativized possessive NPs obligatorily require the resumptive pronouns but others don’t. It depends on the distance of the relationship between the relativized possessor and the possessed NP. If the relationship between the possessor and the possessed is quite close, like a man and his wife as in example (37), the resumptive pronoun is optional. If the relationship between the possessor and the possessed is quite distant, such as between a man and his house as in example (38), the resumptive pronoun is obligatorily present in the relative clause.

(35)

a. dek khít wā phūuchaay tī māa
   child think that man hit dog
   ‘The child thinks that the man hit the dog.’

b. phūuchaay [thī dek khít wā (khāw) tī māa] niŋ pay lēew
   man REL child think that he hit dog run away go already
   ‘The man that the child thinks hit the dog ran away.’

c. māa [thī dek khít wā phūuchaay tī (man)] niŋ pay lēew
   dog REL child think that man hit it run away go already
   ‘The dog that the child thinks the man hit ran away.’

(36)

a. chān rèw wā phūuchaay hāy nōn dek
   I know that man give money child
   ‘I know that the man gives the child some money.’
b.  

\[ \text{dēk} [\text{thī chān rū wā phūuchaay háy ēn (khāw)] pen nákrian}
\]

child REL I know that man give money he be student

‘The child to whom I know the man gives some money is a student.’

(37)

a.  

\[ \text{chān rū wā phanrayaa khōṣ ṭhū phūuchaay pūay}
\]

I know that man of man sick

‘I know that the man’s wife is sick.’

b.  

\[ \text{phūuchaay [thī chān rū wā phanrayaa (khōṣ khāw)] pūay māy maa}
\]

man REL I know that wife of he sick not come

‘The man whose wife I know is sick doesn’t come.’

(38)  

\[ \text{phūuchaay [thī chān yū bān (khōṣ) khāw] pen yāat chān}
\]

man REL I live/stay house of he be relative I

‘The man whose house I live in is my relative.’

5. Conclusion

It is found that the occurrence of pronouns in the relative clause is related to the syntactic-semantic relation of the noun phrase being relativized. In Thai, the resumptive pronoun can occur if the relativized noun phrase is any of subject, direct object, indirect object, or possessor. The occurrence of the resumptive pronoun in Khmer relative clauses seems to be more limited. Resumptive pronouns may occur with relativized subject, indirect object, or possessor, but not with direct object. Unlike Thai, the occurrence of the pronoun with a relativized possessor seems to be obligatory.

The distribution of pronoun retention in Khmer is particularly interesting from a more general theoretical perspective. According to Keenan and Comrie (1977), if pronoun retention is possible for a particular position on the Accessibility Hierarchy, then it must also be possible for all positions lower on the Accessibility Hierarchy. Thai does not violate this generalization, since pronoun retention is possible for all positions. But in Khmer, pronoun retention is possible for subjects, but not for direct objects, although direct objects are lower than subjects on the Accessibility Hierarchy, thus providing a direct counterexample to this particular claim of Keenan and Comrie (1977).

References


INTERACTION OF PHONOLOGY AND SEMANTICS IN TAIWANESE NAMES – AN OPTIMALITY THEORY APPROACH

Chi-hua Hsiao
National Taiwan University
chhsiao3@ms65.hinet.net

Abstract
Past literature on Chinese names were always based on sociolinguistics point of view. This paper, however, aims to take cognitive approach to focus on issues relating to phonology and semantics in Taiwanese’ names, and to study how phonological and semantic factors interact with each other. We adopt Optimality Theory as the framework and propose six constraints. These constraints include: (1) Final characters avoid low tone; (2) Names avoid conflicting context; (3) Final characters avoid siyin; (4) Characters with bad meanings are disfavored; (5) Names that do not specify gender are disfavored; and (6) Names homophonic to bad meanings are disfavored. The result shows that semantic constraints are ranked higher than phonological ones. The ranking also sheds light on how cognitive concepts, such as metaphor and metonymy, are reflected in names.

1. Introduction
Naming is very significant in Taiwan society because names not only reflect social expectations but serve as first expression. They are important of several reasons. First, they stand for the social valuation and expectation for an individual. Parents always name their children by good meaning or wishes they hope children to attain. Second, they reflect the cognition behind people’s naming favorites. Due to the fact that everyone wants his name to be good and pleasant, we believe a large-scale study of names will reflect the cognitive preference of phonology and semantics on names. This paper discusses what kinds of names are preferred, and what cognitive principles lies behind these names.

Many researches about names in Western and Chinese society are from social and cultural point of views; however, few of them discuss issues concerning to phonology and semantics, such as perception, production, connotation and gender difference. We believe that name is far from merely a social ritual or activity; more considerably, it reflects particular mental and cognitive inclination. This paper, therefore, aims to study how phonology and semantics interact with each other in Taiwanese’ names, and how cognition is represented in names.

We adopt Optimality Theory (Prince and Smolensky, 1993) (OT, hereafter) as the theoretical framework. One of the core ideas of OT is that every grammar in language is a system of conflicting forces. These conflicting forces together motivate languages to behave the way they are. OT manages to put these factors, called constraints, in a hierarchical ranking. The ranking excludes undesirable candidates and filters the probably best output. Hence, what OT explains is the probability of language behaviors, not the absolute one. It is the same story for names. No sets of rules can predict the exact preference of names. This is why OT could, and should be applied to
provide insights of the ranking of cognitive constraints in names.

In this paper we propose the following constraints.

(1) Final characters avoid low tone. (* Low tone)
Lakoff and Johnson (1980) has claimed the metaphor “GOOD IS UP; BAD IS DOWN” is rooted in human’s cognition. This oriental metaphor is reflected in almost every aspect of human’s life: emotion, social status, health, conscious, and of course, the acoustic perception. By applying this metaphor to tones, we want to know whether final characters of names prefer high tones, that is, first and second tone.

(2) Names avoid conflicting context. (* Conflicting context)
Xu (1994) argues that tones are more clearly distinguished in compatible contexts than conflicting contexts. Sometimes tones are even twisted in conflicting contexts because the ending and beginning value of two adjacent tones are too diverse. We expect our data to support the idea that names are more prone to compose compatible contexts.

(3) Final characters avoid siyin. (* Siyin)
In Chinese phonology, two types of rhyme are classified according to vowel components—hongyin and siyin. Hongyin is the sound symbolism of wideness and strength while siyin of weakness. We will discuss whether hongyin is more favored, and further look for acoustic or biological explanations.

(4) Characters with bad meanings are disfavored. (* Bad meaning)
Semantics is indispensable to names, especially the connotation of each character. We hypothesize that connotation plays a dominate role in naming, and all characters on names have good connotation.

(5) Names that do not specify gender are disfavored. (* No specific gender)
Gender difference of characters is of importance as well. Traditional society likes names which clearly convey masculine and feminine implication. However, more and more names seem to have characters not specific to gender stereotypes.

(6) Names homophonic to bad meanings are disfavored. (* Homophonic to bad meaning)
While a set of sound has more than one correspondence in form, it is a homophonic pair. Liao (2000) suggests that people avoid names homophonic to bad or common things. Names homophonic to bad meaning will activate negative association. This issue will be discussed to see how sounds and meaning influence each other.

With these constraints in mind, we hope to propose a model that can explain the ranking of names.

2. Literature review
Optimality Theory is a framework that reflects universality. Its central claim is every language differs in the ranking of constraints. Its idea on hierarchical ranking can explain all language differences within one set of universal constraints, and also the mental function of human’s cognition Though OT is developed by phonologists, it is also applied to syntax, computational linguistic, cognitive linguistics and language
acquisition. (Kager 1999, Jason, 1997)

Liao (2000) has systematically studied Taiwan-Chinese personal names, nicknames and English names. From a sociolinguistic view, Liao proposes nineteen rules that dominate the naming process, and one of which about phonology is stated as “Easy to write, recognize and pronounce.” She points out that some parents prefer names which could be pronounced both in Mandarin Chinese and South Min. Other rules regarding to semantics include “Names after a famous or memorable contemporary, historical” or fictional figure” and “Having the name to bear the parent’s expectation about the child.”

The 21 Seiki Kenkyukai (2000) examines Chinese names and concludes that parent’s expectations and wishes are often coded in children’s names. For examples, if they hope children to be successful, they choose characters like “昇”(sheng1), “登”(deng1) and “富”(fu4); if they want children to be healthy, they use “安”(an1), “康”(kang1), and “吉”(ji2).

Wilson (1998) studies the meaning of names in Western Europe. He usefully points out that names are not merely words. Instead, names embed each individual and each family in a historical milieu and reflect a deep continuity in each society. In western society, names are identification of people’s religion, social status and family background.

3. Method and Data
3.1 Method
Statistic analysis is required and crucial to our discussion. Due to our hypotheses with reference to the distribution of tones and rhymes, quantitative evidence is the basis of our interpretation. In addition, we adopt OT to explain the interaction of phonology and semantics. Naming is a complex process and no set of rules could absolutely predict it. Therefore, OT is employed to explain the most probabilistic trend.

3.2 Data
The scope of our data is two-character given names of Taiwanese. The data is randomly selected from Telephone Book of Kaoshing City, with totally 1449 names collected. We first control the surnames as “陳” (chen2) and “林”(lin2), since they are the commonest surnames in Taiwan, and could be reasonably assumed to have more list items than other surnames. Next, we pick up first characters of the given names due to their majority and popularity. The following is the layout:

<table>
<thead>
<tr>
<th>Tone</th>
<th>Characters</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>安(an1), 秋(ciu1), 光(guang1)</td>
<td>349</td>
</tr>
<tr>
<td>Second</td>
<td>明(ming2), 宏(hong2)</td>
<td>389</td>
</tr>
<tr>
<td>Third</td>
<td>永(yong3), 雅(ya3), 敏(min3)</td>
<td>348</td>
</tr>
<tr>
<td>Forth</td>
<td>秀(sie4), 俊(jyun4)</td>
<td>363</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1449</td>
</tr>
</tbody>
</table>

As for the second character of the given names, it depends on what we collect from the Telephone Book. Once the surnames and first characters of the given names are decided, we could consequently put more emphasis on the choice of second characters with
better control.

4. Results and discussion

4.1 Final characters avoid low tone (*Low tone)

Orientational metaphors are systemically structured in humans’ cognition. For example, we have HAPPY IS UP, HAVING CONTROL IS UP, and HIGH STATUS IS UP. These spatial orientations arise because our body and physical environment function in the this way: people erect when healthy and lie down when sick; the level of a substance goes up when something is added to it and drops down when something is taken away from it (Lakoff and Johnson, 1980). All of these metaphors are deeply rooted in human nature’s cognition, and it is prevalent in life experience and around our body. In addition to biological explanations, culture also conveys the identical concept. In Taiwan society, we adopt up-and-down spatial relation to mark good and bad. For instance, we have “jia-shang” (A plus) and “jia-sia” (A minus) to grade students’ assignment; we have idioms like “ren-iou-wang-shang- kan, bu-iou-wang-shang-kan” (People should look up, not look down on.) to refer that people should look up to catch up anything valuable.

The metaphor “GOOD IS UP; BAD IS DOWN” is a concept reflected in almost every aspect of humans life: emotion, social status, heath, conscious, and of course, the acoustic perception. Sound is often regarded as an indicator of a person’s health and spirit. If a person pronounces high, loud and clear sounds, people are prone to think he is a high-spirited status; on the other hand, if a person speaks in low-pitched sounds, people think that he is in a bad mood or occupied by something else his mind. Chinese has four tones. According to Chao (1980), the first tone is a high-high tone, marked as 55; the second tone is a mid-high tone, marked as 35; the third tone is mid-low-high tone, marked as 214; and the forth tone is a high-low tone, marked as 51. These numbers stand for the beginning and ending value of each sound, estimated by average sound levels. Ending at the highest value 5, the first and second tones are categorized as high tone; ending at the lowest value 1, the fourth are categorized as low tone. As for the third tone, although Chao claims its pattern as 214, it is actually uttered as 213 or 212 (Fon, 1997). No matter the ending tone is 2 or 3, it belongs to low tone rather than high tone.

Based on the1449 names collected, the distribution of tones are as follows:

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>First tone</td>
<td>397</td>
</tr>
<tr>
<td>Second tone</td>
<td>543</td>
</tr>
<tr>
<td>Third tone</td>
<td>121</td>
</tr>
<tr>
<td>Fourth tone</td>
<td>388</td>
</tr>
<tr>
<td>Total</td>
<td>1449</td>
</tr>
</tbody>
</table>

The pattern for high and low tones is:

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>High tones</td>
<td>940</td>
</tr>
<tr>
<td>Low tones</td>
<td>509</td>
</tr>
<tr>
<td>Total</td>
<td>1449</td>
</tr>
</tbody>
</table>
Roughly 65% of names have high tone as final characters. This means that people unconsciously tend to choose characters with high tones for their children’s names. Especially the second tone, a rising tone, is particularly prevalent as the final character. We suggest the reason may be that in terms of sound production, high tones are easier to pronounce while we call out someone’s name, and rising tones associate names and peoples in a good situation. The tendency for high tone in final characters demonstrates that the “GOOD IS UP” metaphor is reflected by human’s auditory sense.

4.2. Names avoid conflicting context (*Conflicting context)
When characters are conjoined together, the tones would influence each other, especially in running speech. Xu (1994) studies coarticulated tones and finds that the articulation of tones depends on the nature of tonal context. He distinguish two types of contexts: (1) a compatible context is an environment in which adjacent characters have similar values along the phonetic dimension, and (2) a conflicting context is an environment in which adjacent characters have very different values along the phonetic dimension. Xu suggest that in a compatible context, the deviation is relatively minor and tones are basically maintained, while in a conflicting context, the deviation is much greater and tones are easily distorted. As for perception, tones are easier to be identified in compatible context and more difficult in conflicting contexts.

There are totally sixteen tone combinations. Because Xu does not clearly define under what condition a compatible and conflicting context will be composed, we decide that once the value contrast is more than one level, a conflicting context is made.

<table>
<thead>
<tr>
<th>Type</th>
<th>Adjacent Tones</th>
<th>Type</th>
<th>Adjacent Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compatible context</td>
<td>1 (55) 1 (55)</td>
<td>Conflicting context</td>
<td>1 (55) 2 (35)</td>
</tr>
<tr>
<td></td>
<td>1 (55) 4 (51)</td>
<td></td>
<td>1 (55) 3 (214)</td>
</tr>
<tr>
<td></td>
<td>2 (35) 1 (55)</td>
<td></td>
<td>2 (35) 2 (35)</td>
</tr>
<tr>
<td></td>
<td>2 (35) 4 (51)</td>
<td></td>
<td>2 (34) 3 (214)</td>
</tr>
<tr>
<td></td>
<td>3 (212) 2 (35)</td>
<td></td>
<td>3 (212) 1 (55)</td>
</tr>
<tr>
<td></td>
<td>3 (212) 3 (214)</td>
<td></td>
<td>3 (212) 4 (53)</td>
</tr>
<tr>
<td></td>
<td>4 (53) 2 (35)</td>
<td></td>
<td>4 (53) 1 (55)</td>
</tr>
<tr>
<td></td>
<td>4 (53) 3 (214)</td>
<td></td>
<td>4 (53) 4 (51)</td>
</tr>
</tbody>
</table>

1 To demonstrate first tone and second tone do not necessarily have prevalence in other genre, we have examined 1186 Chinese characters by random selection. We found that the distribution of four tones is as follows: 20.4% for first tone, 25% for second tone, 20.3% for third tone and 34.3% for fourth tone. This examination shows that first tone and second tone don’t have an preponderance as they do in names.

2 According to Fon (1997), the third is realized as 212 or 213 when followed by another tone. Here we follow Fon’s study because that is the real realization in oral conversation.

3 According to Fon (1997), the fourth tone is realized as 53 rather than 51 when followed by another tone.
In the following, we will discuss each tone by details.

4.2.1 First tone as first character
We have controlled the first character of given names as “安”(an1), “秋”(ciu1) and “光”(guang1), and totally 349 tokens are collected. The tonal distribution of the second character is as follows:

<table>
<thead>
<tr>
<th>Tone of following characters</th>
<th>Tokens</th>
<th>Percentages</th>
<th>Context Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>89</td>
<td>25.5 %</td>
<td>Compatible</td>
</tr>
<tr>
<td>Second</td>
<td>143</td>
<td>41 %</td>
<td>Conflicting</td>
</tr>
<tr>
<td>Third</td>
<td>34</td>
<td>9.7 %</td>
<td>Conflicting</td>
</tr>
<tr>
<td>Fourth</td>
<td>83</td>
<td>23.8 %</td>
<td>Compatible</td>
</tr>
</tbody>
</table>

Compatible context of character combinations occupies 49.3% and the conflicting context 50.7%. This means first tone shows no preference for compatible or conflicting tonal context. What is noteworthy is that second tone has a dominate preference, even though it leads to a conflicting context with first tone.

4.2.2 Second tone as first character
We have set first character of given names as “明”(ming2) and “宏”(hong2), with totally 389 tokens collected. The tonal distribution of the second character is:

<table>
<thead>
<tr>
<th>Tone of following characters</th>
<th>Tokens</th>
<th>Percentages</th>
<th>Context Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>111</td>
<td>28.5 %</td>
<td>Compatible</td>
</tr>
<tr>
<td>Second</td>
<td>122</td>
<td>31.4 %</td>
<td>Conflicting</td>
</tr>
<tr>
<td>Third</td>
<td>34</td>
<td>8.7 %</td>
<td>Conflicting</td>
</tr>
<tr>
<td>Fourth</td>
<td>122</td>
<td>31.4 %</td>
<td>Compatible</td>
</tr>
</tbody>
</table>

The result is that compatible context occupies 59.9%, roughly 60% of character combinations. The situation is slightly different from the case of first tone as first character. The chance of each tone to follow second tone seems more equal (28.5%, 31.4%, 31.4%), except for third tone, a tone more difficult to pronounce in nature. It is reasonable to assume that second tone is more acceptable to any other tones while the preference for compatible context also maintained.

4.2.3 Third tone as first character
We have set the first character of given names as “永”(yong3), “雅”(ya3) and “敏”(min3), with 348 tokens collected. The tonal distribution of the second character is:
Phonology and semantics in Taiwanese names

Table 6

<table>
<thead>
<tr>
<th>Tone of following characters</th>
<th>Tokens</th>
<th>Percentages</th>
<th>Context Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>永(yong3), 雅(ya3), 敏(min3)</td>
<td>First</td>
<td>102</td>
<td>29.3 %</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>126</td>
<td>36.2 %</td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>19</td>
<td>5.5 %</td>
</tr>
<tr>
<td></td>
<td>Fourth</td>
<td>101</td>
<td>29 %</td>
</tr>
</tbody>
</table>

Conflicting context occupies the majority of tonal coarticulation (58.3%). Although second tone has the highest frequency 36.2%, it is still overwhelmed by first and forth tone, with which conflicting context is produced.

4.2.4 Forth tone as first character
We have set the first character of given names as “秀”(sie4) and “俊”(yjun4), with 363 tokens collected. The tonal distribution of the second character is:

Table 7

<table>
<thead>
<tr>
<th>Tone of following characters</th>
<th>Tokens</th>
<th>Percentages</th>
<th>Context Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>sie, yjun</td>
<td>First</td>
<td>95</td>
<td>26.2 %</td>
</tr>
<tr>
<td>秀, 俊</td>
<td>Second</td>
<td>152</td>
<td>41.9 %</td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>34</td>
<td>9.4 %</td>
</tr>
<tr>
<td></td>
<td>Fourth</td>
<td>82</td>
<td>22.5 %</td>
</tr>
</tbody>
</table>

In the 363 tokens data, compatible context has 51.3%. This figure is not significant since there is only half chance for compatible context. Although second tone again has the largest proportion of distribution, it is still outnumbered by first and fourth tone.

4.2.5 Overview
By examining all the data, we have the final analysis of coarticulated tonal context for given names:

Table 8

<table>
<thead>
<tr>
<th>Token</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compatible context</td>
<td>736</td>
</tr>
<tr>
<td>Conflicting context</td>
<td>713</td>
</tr>
</tbody>
</table>

This quantitative analysis reveals that people have no preference for compatible tonal context, because the chance for both two tonal contexts is roughly parallel. This implies that people don’t tend to make names concerning about the tonal pattern. There must be some factors that override this consideration. In the following, we are going to focus on the semantics of characters. Before that, let’s review the intermit summary so far:
High tones (first and second tones) do have priority to be final tones, especially second tone, which has 37.6% of distribution among all.

The distribution of each tone is average:
- First tone has 25.5%, 28.5%, 29.3% and 26.2% chance to follow first, second, third and fourth tone, respectively;
- Second tone has 41%, 31.4%, 36.2% and 41.9% chance to follow first, second, third and fourth tone, respectively;
- Third tone has 9.7%, 8.7%, 5.5% and 9.4% chances to follow first, second, third and fourth tone, respectively;
- Fourth tone has 23.8%, 31.4%, 29% and 22.5% chance to follow first, second, third and fourth tone, respectively.

This means that the first character of given names does not particular prefer any tone to follow it. That no matter what tone the first character is, the second tone always has the highest percentage is an explicit indicator.

Each tone has its lowest proportion while the proceeding character carries the same tone with it. We suggest the reason lie in that people favor tonal change in adjacent units. In this way sounds are more melodic and cadent.

There is a no tendency for names to comprise compatible context or conflicting context. That the second tone always has the highest proportion in all environment is again an obvious indicator.

4.3. Final characters avoid siyin (*Siyin)

Each Chinese character has three parts: onset, rhyme, and pitch. For rhymes, they are categorized to two groups: hongyin and siyin, depending on the glides of rhymes. Rhymes without glides like [pa] (怕), or rhymes with glide [u] like [uai] (外) are classified as hongyin; rhymes with glide [i] like [ie] (也), or glide [y] like [yan] (全) are classified as siyin. The classification is based on phonetic feature. Siyin have [+high] and [-back] features. Hongyin are sounds that are louder and clearer than siyin because of the manner of articulation and place of articulation. Our data shows that hongyin has a greater chance to be adopted in final characters:

<table>
<thead>
<tr>
<th></th>
<th>Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hongyin</td>
<td>842</td>
<td>58.1%</td>
</tr>
<tr>
<td>Siyin</td>
<td>607</td>
<td>41.9%</td>
</tr>
<tr>
<td>Total</td>
<td>1449</td>
<td>100%</td>
</tr>
</tbody>
</table>

Although the difference is not very significant, our data still proves a preference for hongyin rather than siyin.

4.4 Characters with bad meanings are disfavored (*Bad meaning)

Connotation of each character is a crucial consideration for naming. Like we have just discussed, although phonological pattern is important, semantics seems to be a more powerful motivation. According to our analysis, the tonal context of names does not show any preference for compatible context; hence, we would turn to semantics to survey other considerations.

First, when the first character is a first tone, second or third tones would
Phonology and semantics in Taiwanese names

comprise a conflicting context. Characters with second or third tone in our data include: “元”(yuan2), “平”(ping2), “宏”(hong2), “男”(nan2), “成”(cheng2), “祥”(siang2), “妮”(ni2), “如”(ru2), “池”(chih2), “文”(wun2), “美”(mei3), “眼”(ian3), “景”(jing3) and many others. Obviously, these characters either have positive meaning or mark gender difference. Take the following characters “成”(cheng), “美”(mei) and “妮”(ni) for example. The meaning of “成”(cheng) is “result” and “success”; that of “美”(mei) is “beauty” and “good”; that of “妮”(ni) is “young” and “beauty”. They are chosen because of their good meanings. The conflicting context, on the other hand, seems to be ignored.


What is worthy of notice is, several characters in our data, including “矗”(chu4), “珍”(chen1), “圳”(jung4) and “釗”(jhao1), are difficult to read, due to their low frequency and low popularity; however, they are still chosen as part of names even though they constitute a conflicting tones. Not surprisingly, they all have good or neutral meaning. “矗”(chu4) means “highness”; “珍”(chen1) means “values”; “圳”(jung4) means “trench”; “釗”(jhao1) means “diligence”. These words prove that once a character bears good meaning, the phonological factors become secondary.

Sometimes tonal context conflicts with semantic connotation. By examining names which compose a conflicting context, we find that connotation of each character has priority over tonal context consideration. If a character bears positive meaning, it would be adopted even though a conflicting tone is resulted. Louise (1998) also points that many Chinese people check the dictionary to name their children to make sure the character is resourceful and have good meaning. No matter by what means parents pick up characters, good meaning is definitely a very dominate factor.
4.5 Names that do not specify gender are disfavored (No specific gender)
Many Chinese characters imply gender stereotypes. For example, people feel “成” (cheng2) “榮” (rong2) and “淵” (yuan1) are for male; “玲” (ling2), “靜” (jing4) and “蓉” (rong2) are for female. Culture and social expectation associate these characters with specific gender implications. The meaning of “成” (cheng2) is “success” and “result”; “榮” (rong2) is “honor” and “good reputation”; “淵” (yuan1) is “knowledgeable”. Both men and women could achieve these goals; however, the society tends to reserve these characters for male, not female. This is the same for feminine characters. Men deserve positive meanings like “玲” (ling2) for “favor”, “靜” (jing4) for “stillness” and “peace”, “蓉” (rong2) for “good-looking”, but the society is prone to apply these meanings to women. In our 1449 tokens, we do find that all names have either two characters or at least one character for specific gender. For feminine meanings, there are “姿” (zih1), “娜” (na4), “莉” (li4), “琪” (ci2), “麗” (li4), “香” (xiang1), “婷” (ting2), “梅” (mei2) and so on; for masculine meaning there are “宏” (hong2), “泉” (cyuan2), “光” (guang1), “俊” (yun4), “建” (jian4), “郎” (lang2) and so on; for neutral meanings there are “喜” (si3), “恩” (en1), “智” (jhih4), “旻” (min3), “穎” (ying3), “德” (de3) and “庭” (ting2). Many names have both their characters for masculine or feminine meaning. For example, we have “宏光” (hong2 guang1), “建俊” (jian4 jyun5), “姿婷” (zih1 ting2) or “莉香” (li4 xiang1). As for names which have one character for neutral meaning, the other character must carry gender implication. For example, “喜榮” (si3 rong2) or “成穎” (cheng2 ying3). “喜” (si3) and “穎” (ying3) are neutral characters, but the other characters, “榮” (rong2) and “成” (cheng2), help to identify these names for men. Although neutral characters do not tell gender, their counterparts take this job. As the society becoming more liberal and more neutral characters chosen as names, the gender stereotype is still influential, because it presents the social and cultural expectation children should attain.

4.6 Names homophonic to bad meanings are disfavored. (Homophonic to bad meaning)
Liao (2000) proposes that people avoid names homophonic to bad or common things. Homophonic pairs are not unusual in Chinese, since a set sound may have several correspondences in forms. Liao gives an example of “湯曉歆” (tang1 siao3 sin1). The person with this name is always upset when being warned to be careful on drinking soup. This illustration explains why people try to avoid homophonic names which evoke inadequate meanings. In our opinion, homophonic sound is based on metonymic association. That is, sound is part of the mane; sound with positive association leads to a positive impression of name, and sound with negative association leads to a negative one. In our data, we have homophonic names, but all of them bear good, or at least neutral, associations: “明裕” (ming2 yu4) is linked with “good reputation”; “明娥” (ming2 e2) with “quota”; “雅觀” (ya3 huai1) with “good looking”; and “敏瑞” (min3 rui4) with “sharpness”. It is reasonable to assume that names homophonic to bad meanings are abandoned. Consequently, our data does not have such example. However, names like “建仁” (jian4 ren2), “建銘” (jian4 ming2) and “健民” (jian4 min2) are popular in Taiwan, and they are homophonic to “lowly people” which is an insult or personal attack for swearing. Why do parents adopt these names? The separate meaning of “健” (jian4), “仁” (ren2), “銘” (ming2) and “民” (min2) is good, which means “strong”, “kindness”, “forever” and “people” respectively. These good meanings are the reasons
Phonology and semantics in Taiwanese names

why these names are prevalent, even though a bad association arises. Again, connotation of each character is dominating. The ranking of connotation is more prominent than the homophonic constraint.

5. Interaction between Phonology and Semantics

In the previous sections, we have discussed phonological and semantic constraints of names. We employ OT to provide a model to explain the complexity of names:

(1) Final characters avoid low tone. (*Low tone)
(2) Names avoid conflicting context. (*Conflicting context)
(3) Final characters avoid siyin. (*Siyin)
(4) Characters with bad meanings are disfavored. (*Bad meaning)
(5) Names that do not specify gender are disfavored. (*No specific gender)
(6) Names homophonic to bad meanings are disfavored. (*Homophonic to bad meaning)

Let's first consider the ranking of (1) and (4). Constraint (1) states that final characters avoid low tone while (4) states that characters avoid bad meaning. In our data, all characters conform to (4) while not all final characters avoid low tones. Examples are: “安道” (an1 dao4), “秋彥” (ciu1 yan4), “光宇” (guang1 yu3), “明有” (ming2 you3), “永郁” (yomg3 yu4), “秀智” (sie4 jhih4) and “雅莉” (ia3 li4) are many among the others. Table 2 represents that roughly 35% names in our data violate constraint (1) because they end at low tone. Under the situation that all names conforms to (4) but 35% violate (1), the ranking is (4) > (1).

As for constraint (4) and (6), they sometimes conflict with each other. The former says characters avoid bad meaning and the latter says the name should not be homophonic to bad implication. In our data, these two constrains are both satisfied because all characters have good meanings and all homophonic names are connected with either good or neural associations. Based on our data, the ranking is equally important. However, if we take names like “健仁” (jian4 ran2), “健民” (jian4 min2) and “健銘” (jian4 ming2) into consideration, constraint (4) is ranked higher than (1). These names demonstrate that the violation of (6) is also acceptable. Therefore, constraint (4) > (6) is possible as well.

Regarding to constraint (1) and (6), if we only consider data collected in this study, all names are subject to constraint (6) while 35% of names violate (1). It is reasonable to propose (6) > (1), because our data shows no names homophonic to bad meanings. On the other hand, if we take cases of “健仁” (jian4 ran2), “健民” (jian4 min2) and “健銘” (jian4 ming2) into consideration again, the situation is different. These three names happen to end at high tone; therefore constraint (1) is ranked over (6). Hence: (1) > (6) is also a possible ranking.

As for constraint (2) which states names avoid conflicting context, Table 8 displays that this trend is not obvious. What’s more important is the connotation of characters and high tones. Hence, constraint (2) is ranked lower than (1), (4), (6).

As for constraint (1) and (2), 35% of names violate (1) while roughly 50% of names violate (2). This statistic analysis implies that more data is subject to (1) than (2). Table 4 to Table 7 also clearly elucidates that second tone (a high tone) has the largest proportion no matter what tone proceeds it. Namely, if the proceeding character tone is
first or second tone with which a conflicting context will be produced, second tone is still the one that has the highest frequency. Example are “安達”(an1 da2), “安吉”(an1 ji2), “秋伶”(ciu1 ling2), “秋怡”(ciu1 yi2), “明光”(ming2 guang1), “明學”(ming2 syue2), “宏青”(hong2 cing1) and “宏潔”(hong2 jie2) and many others. Hence, constraint (1) is ranked higher than (2).

As for constraint (3) which requires final characters avoid siyin, Figure 9 shows that 58.1% of names conforms to this constraint. If we consider it with constraint (2) together, we find that (3) is ranked higher than (2). The statistic evidence shows more names satisfy constraint (3) (58.1%) than (2) (49.5%). Names of this case include: “安宏”(an1 hong2), “秋文”(ciu1 wun2), “光此”(guang1 cih3), “明哲”(ming2 jhih2), “永兆”(yong3 jhao4), “雅珍”(ia3 jhen1), “敏志”(min3 jhih4), “秀芳”(sie4 fang1) and “俊富”(jyun4 fu4) and many others. However, (3) is still ranked inferiorly than (1), (4) and (6) because these three constrains are supported by the majority of cases.

As for (5) which states that names not specifying gender are disfavored is of significance, because our data all implies gender difference, either by two characters or only one character. For names having one neutral character, the other character must tell the gender implication. Examples are “秋彥”(ciu1 yan4) and “秋伶”(ciu1 ling2). The character “秋”(ciu1) does not carry gender stereotype; however, “彥”(yan4) is obviously for male and “伶”(ling2) for female, with accordance to the definition and social expectation of these two characters respectively. Constraint (5), therefore, overrides (1), (2) and (3).

As for constraint (4) and (5), they are equally important because all data complies them. We can see that in names, semantics consideration is more powerful than phonological ones. No matter how phonological constraints (1), (2) and (3) are ranked, semantic ones have priority order.

Based on our discussion, two possible ranking are proposed:

<table>
<thead>
<tr>
<th>(1)</th>
<th>*No specific gender</th>
<th>*Bad meaning</th>
<th>*Homophonic to bad meaning</th>
<th>*Low tone</th>
<th>*Siyin</th>
<th>*Conflicting context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidate b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This ranking explains most data. We take several names to examine the ranking⁴:

<table>
<thead>
<tr>
<th>(4)</th>
<th>* No specific gender</th>
<th>* Bad meaning</th>
<th>* Homophonic to bad meaning</th>
<th>* Low tone</th>
<th>* Siyin</th>
<th>* Conflicting context</th>
</tr>
</thead>
<tbody>
<tr>
<td>光博 (guang1 bo2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

⁴ OT has a fallacy of perfection. That is, no output form is possible to satisfy all constraints. In our study, some names satisfy all constraints. The reason is that our constraints do not include all possible ones in naming. Say for example, fortune teller’s suggestions or family’s pedigree names are beyond our discussion.
If we take common names like “健仁” (jian4 ran2), “健民” (jian4 min2) and “健銘” (jian4 ming2) into consideration, we have the second version of ranking. Although these names are not collected in our data, they are still discussed due to the commonplace in Taiwan.

(2)

<table>
<thead>
<tr>
<th>Candidate a</th>
<th>Candidate b</th>
</tr>
</thead>
<tbody>
<tr>
<td>* No specific gender</td>
<td>* Bad meaning</td>
</tr>
<tr>
<td>* Low tone</td>
<td>* Homophonic to bad meaning</td>
</tr>
<tr>
<td>* Siyin</td>
<td>* Conflicting context</td>
</tr>
</tbody>
</table>

This second possible mode is proposed based on the cases of “健仁” (jian4 ran2), “健民” (jian4 min2) and “健銘” (jian4 ming2). The following is the illustration:
The three names demonstrate constraint (6) ‘Names homophonic to bad meanings are disfavored.’ is not fatal. Since the three name are common in Taiwan society, this ranking is possible, although is not reflected by our data.

6. Future study
This paper focuses on linguistics, that is, phonology and semantics of names. We provide an OT model to explain the most probable constraint ranking of names. Further work can be extended to culture and personal considerations, such as family background and personal preference. Many Taiwanese name their children by fortune tellers’ suggestions, specific characters of family generations, and convenience of pronunciation of their native languages. The large-scale interview is a potential method to survey these social factors. Once these factors are included, OT could provide a more complete model of names.

References
Phonology and semantics in Taiwanese names

Society of America, 95: 2240-53.
二十一世紀研究會，《人名的世界地圖》，時報文化，2002。
吉常宏著，《中國人的名字別號》，商務印書館，1994。
趙元任著、丁新邦譯，《中國話的文法》，香港中文大學，1968，中譯本1980。
董同龢著，《漢語音韻學》，文史哲出版社，1998。
THE TONE AND SYLLABLE STRUCTURE OF TURUNG

Stephen Morey
Research Centre for Linguistic Typology
La Trobe University
moreystephen@hotmail.com

1. Introduction – the Turung

The Turung, also called Tai Turung or Tairong, are a group of around 1200 people living in seven villages in Assam, India. Four of these villages are in Jorhat district, two in Golaghat district, and one in the Karbi Anglong autonomous district, all of which are on the south side of the Brahmaputra River.

Today the Turungs speak a language that they call Turung or Singpho or sometimes simply ʔiiʔ buuʔ gaʔ ‘our language’. This language is part of the Jingpho group within Tibeto-Burman, but its lexicon contains a significant Tai admixture, and Turung differs in terms of both phonology and syntax from Jingpho as has been described by Hanson (1896), Matisoff (1974a) and Dai (1992). Some dialects of Singpho spoken in India (one form of which was described by Needham 1889) are not intelligible to Turung speakers, and there appears to be a dialect continuum between Turung and the various Singpho varieties.

The Turungs believe that they were once a Tai speaking people, and that their name was originally ʔaiʔA4 loŋA1, a Tai term meaning ‘great Tai’. British sources dating back to the early 19th century always describe the Turungs as a Tai speaking group. For example, Grierson (1904) described them as one section of the Shans (i.e. Tais) “who at various times entered Assam has retained this name, and its members are now known as the Tairongs, Turūngs or Shām (i.e. Shān) Turūngs.”

Grierson added that at the time that he was writing, there were very speakers of Turung/Tairong, putting the number at 150. The information he gives about the language shows that it was Tai, albeit slightly aberrant (see Morey 2005:44). The reason for the small number people speaking Turung/Tairong as a Tai language was that this group had been taken prisoner by the Kachins (or Singphos) and that “during their servitude to the Kachins they entirely forgot their own language, and now only speak that of their conquerors, Singpho.” (Grierson 1904:167).

---

1 This paper is a much revised version of one presented at the SEALS XV meeting in Canberra. I acknowledge the helpful comments of those who attended on that occasion, and also Alexandra Y. Aikhenvald, Alec Coupe, Randy LaPolla, Mark Post and Marija Tabain. I am particularly grateful to Kon Kham Turung, of Rengmai village, who was my main consultant for the data included in this paper. Abbreviations in this paper are as follows:

- A.AG anti-agentive
- C consonant
- DEF definite
- GV generalised verb
- N nasal
- POSS possessive
- PRT particle
- T tone
- V vowel

2 Also spelled Jinghpaw and Jinghpo (e.g. in Dai and Diehl 2003). The spelling adopted here, also used by Matisoff (2003), parallels the spelling Singpho used in India. In Burma, the language is also called Kachin, a Burmese word referring to various Tibeto-Burman speaking groups in Kachin state, of whom the Jingpho are the most numerous (see Leach 1964 for a fuller discussion).
Table 1: *Aiton and Turung villages*

<table>
<thead>
<tr>
<th>Description of village</th>
<th>Examples</th>
<th>Symbol on map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aiton speaking villages (Turung women are married into the village, but their children generally do not learn Turung)</td>
<td>Duburoni (Ban Nam Thum), Bargaon (Ban Lung)</td>
<td>♦</td>
</tr>
<tr>
<td>Aiton speaking villages with minority Turung speaking households</td>
<td>Tengani, Ahomani</td>
<td>★</td>
</tr>
<tr>
<td>Previously Turung speaking village that has switched to Aiton speaking</td>
<td>Barhula</td>
<td>●</td>
</tr>
<tr>
<td>Village with both Aiton speaking and Turung speaking areas of roughly equal size</td>
<td>Balipathar</td>
<td>□</td>
</tr>
<tr>
<td>Turung speaking village where the language shows a high level of Tai mixing</td>
<td>Rengmai</td>
<td>X</td>
</tr>
<tr>
<td>Turung speaking village (Aiton women are married into the village, but their children do not learn Aiton)</td>
<td>Basapathar</td>
<td>♥</td>
</tr>
</tbody>
</table>

Map of the middle Dhonsiri River, south of the junction of the Doiyang and the Dhonsiri; scale bar – about 10km
The tone and syllable structure of Turung

The Turungs themselves do not accept that they were enslaved by Kachins or Singphos, but their own histories do state that they lived in close proximity to the Singphos in Upper Assam, and hence acquired Singpho language. Modern Singpho scholars from Upper Assam, on the other hand, regard the Turung as one of the clans (khong) of the Singpho who got their name because they lived in the neighbourhood of the Tarung River, which is in Kachin State and the Sagaing Division of northern Burma.

Many older Turungs maintain that their parents and grandparents spoke Tai, and that the Turung community was once Tai speaking, whereas the present-day Turungs are a mixed community. The Turung caste or family names, which represent the male line in Turung villages, include both Singpho forms like la gung\(^1\) and Tai forms like man\(^3\) nuu\(^3\), from maan\(^3\) nuu\(^1\), which in Tai Aiton means ‘village-up’.

Since making an agreement with the Tai speaking Aitons in the early 19\(^{th}\) century, the Turungs have intermarried with Aitons. In all Turung villages, some Turungs have married Aiton wives, and many Turung women have moved to Aiton speaking villages and married Aitons. This has led to a complex contact situation between Aiton and Turung. In the Dhonsiri valley, straddling the borders of Karbi Anglong and Golaghat districts, there are 5 Aiton villages and 3 Turung villages as shown on the map below. Table 1 sets out the linguistic situation in each of the villages:

All Turungs speak Assamese, the language of wider communication in Assam. Some also speak English, and some can speak Hindi. At least partial passive knowledge of these two national languages of India is widespread. In addition, some Turungs have competence in the languages of other tribal groups with whom they are in contact.

2. The segmental phonemes of Turung

The consonant inventory of Turung is presented in Table 2:

<table>
<thead>
<tr>
<th>Consonant Phonemes in Turung</th>
<th>Bilabial</th>
<th>Dental / Alveolar</th>
<th>Palatal / affricate</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless unaspirated stops</td>
<td>p</td>
<td>t</td>
<td>c ([tc])</td>
<td>k</td>
<td>?</td>
</tr>
<tr>
<td>Voiceless aspirated stops</td>
<td>ph</td>
<td>th</td>
<td></td>
<td>kh</td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>b</td>
<td>d</td>
<td>j ([dz])</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
<td>n</td>
<td>ny</td>
<td>ng</td>
<td></td>
</tr>
<tr>
<td>Voiceless fricative</td>
<td>s</td>
<td></td>
<td></td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>Semi vowel</td>
<td>w</td>
<td></td>
<td>y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhotic Approximant</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral Approximant</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This inventory differs significantly from the syllable onset inventory given for Jingpho (Dai and Diehl 2003:402) in several ways. Some of the differences reflect differences of analysis or presentation: Dai and Diehl include initial stop-glide sequences in their table, which we treat as clusters, and they use IPA symbols where we use digraphs. Some of the differences, however, are significant. Turung has an /h/
Stephen Morey

phoneme, which is found only marginally in Jingpho\(^3\). This phoneme is found with words borrowed from Assamese or Tai, but does occur with some native words such as hong\(^2\) ‘sunlight’, apparently a reflex of proto Tibeto-Burman *hwaj ‘shine, bright, yellow’ (reconstruction in Matisoff 2003:430).

Only one other fricative, /s/, is found in Turung (although [f] or [ɸ] are variants of /ph/ especially in younger person’s speech). Jingpho in Burma makes a distinction between the initial of ‘to eat’ <Sha> and ‘to go’ <Sa> (Hanson 1906), but both these words have the same initial in Turung, although with different tones. Matisoff (1974b) gives the Jingpho words for ‘die’ and ‘ten’ as sī and sī respectively, whereas in Turung both words are identically pronounced sī\(^3\).

Permitted final consonants in Turung are listed in Table 3:

**Table 3: Final Consonant Phonemes in Turung**

<table>
<thead>
<tr>
<th>Bilabial</th>
<th>Dental/Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless unaspirated stops</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td>?</td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td></td>
</tr>
</tbody>
</table>

Final -k is uncommon with proto Tibeto-Burman final *-k having regularly become -ʔ in Turung, as it has in Jingpho (see Matisoff 2003:319). Most words with final -k are Tai or other borrowings, but at least one very common Turung function word, bok\(^3\) ‘all’, also has final -k.

Phonetically both /l/ and /r/ are sometimes found syllable finally in Turung. This comes about as follows: the word dai\(^3\) la khan\(^3\) ‘therefore’ is a compound formed from three words: dai\(^3\) naa\(^3\) khan\(^3\) ‘that-POSS-cause’. In fast pronunciation, the schwa in the sesquisyllabic structure reduces, and the word can be realised as [dail khan].

The Turung vowels are presented in Table 4:

**Table 4: Vowels in Turung**

<table>
<thead>
<tr>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

Vowel length is not contrastive in Jingpho, but it is in Turung, at least marginally. For example, peen\(^3\) ‘to fly’ clearly has a long vowel (0.30\(^\prime\)), contrasting with naa\(^3\) pen\(^3\) ‘earlobe’ that has a short vowel (0.19\(^\prime\)). Words with nasal codas and long [e] or long [a] in Turung are cognate with words that are realised in Jingpho with initial stop-glide clusters (e.g. pyen ‘to fly’). The reflex of this glide is expressed as length in Turung. The length of the vowel in peen\(^3\) is the same as the length of the vowel when a word has no consonantal coda, so that words like phee\(^3\) ‘anti-agentive’ are also written in this paper with long vowels.

---

\(^3\) Hanson (1906) lists four words with initial /h/ that are interjections, borrowings or alternative pronunciations of function words. In the Jingpho Miwa Ga Ginsi Chyum, there are seven words with initial /h/, mostly proper names or interjections.
The tone and syllable structure of Turung

The schwa is found as the vowel on the minor syllable of a sesquisyllabic structure (see below Section 4.3).

3. The tones of Turung – a brief survey
A minority of the Turung lexicon, nevertheless containing some of the most basic vocabulary, is monosyllabic. In September 2003, Kon Kham Turung, a 20 year old Turung/Assamese/English trilingual speaker of Rengmai village, suggested that for monosyllables, Turung had five tones, three on syllables with sonorant finals and two on syllables with stop finals. He exemplified these as shown in Table 5:

<table>
<thead>
<tr>
<th>Tone No.</th>
<th>Description of tone</th>
<th>Pitch contour (Hz)</th>
<th>example</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>Low falling</td>
<td>135-110</td>
<td>waa₁</td>
<td>‘DEF’</td>
</tr>
<tr>
<td>Tone 2</td>
<td>Rising then high falling</td>
<td>120-170-90</td>
<td>waa₂</td>
<td>‘return’</td>
</tr>
<tr>
<td>Tone 3</td>
<td>Level</td>
<td>140-120</td>
<td>waa₃</td>
<td>‘tooth’</td>
</tr>
<tr>
<td>Tone 4</td>
<td>Low falling and short</td>
<td>150-120</td>
<td>wa⁴</td>
<td>‘bite’</td>
</tr>
<tr>
<td>Tone 5</td>
<td>Rising then high falling and short</td>
<td>125-155-140</td>
<td>wa⁵</td>
<td>‘pork’</td>
</tr>
</tbody>
</table>

When this was discussed with other Turung informants, it was felt that there was a third tone on stopped syllables, a level tone as in wa⁴, ‘make (a basket)’ a word that Kon Kham Turung maintained had the same tone as wa², ‘bite’.

Since the profile of the two stopped tones is similar to tones 1 and 2, three tones are posited for Turung monosyllables, exemplified in (1):

1) 1 Falling         waa₁, ‘DEF’;  wa², ‘bite’
2 High rising then falling waa₂, ‘return’;  wa⁴, ‘pig’
3 Level             waa₃, ‘tooth’;  wa⁵, ‘make (a basket)’

Subsequent to the discussion leading to the analysis presented in Table 5, a word list of about 1200 items was recorded by Kon Kham Turung, and this list will form the basis of the analysis presented here. The pitch range as measured in Hz that Kon Kham Turung exhibited when recording the original five tones was somewhat lower than the range he used when recording the larger word list that forms the basis for all the other analysis in this paper.

---

4 This was recorded on a Sony Digital Handycam TRV110E Camcorder with a stereo microphone.
5 This list was recorded using a backup Sony Cassette player, with stereo microphone. At the time that the word list was being recorded, neither my handycam nor minidisc player were functioning.
Stephen Morey

Using the Speech Analyzer Version 1.5, the contours of all monosyllabic tones were closely examined, and a further tone shape was observed, a very steep falling tone, as found on the word sai² ‘blood’. This tone was observed only when the initial consonant was voiceless and thus we suggest that it is an allotone of the rising-falling tone (tone 2 in Table 5), although native speakers who were consulted did not always agree that these tones were the same. In (2) we compare two Turung words with this tone, where the pitch of the tone has been given in Hz.

2) sai² ‘blood’ (pitch range 200-120 Hz)
   dai² ‘father’s younger brother’ (pitch range 130-180-110 Hz)

When the initial is voiceless, the pitch of the highest point of the tone is higher than when the initial is voiced, as shown in the following diagram, produced in Speech Analyzer, comparing nam² ‘forest’ and phun² ‘wood’.

Table 6: Pitch chart comparing nam² ‘forest’ and phun² ‘wood’.

The tone marked with the number 1 also has a falling contour, but falls more gently and from a lower starting position, as shown in (3)

3) jon¹ ‘mongoose’ (pitch range 150-120 Hz)

The third tone is level. Analysis with Speech Analyzer sometimes shows pitch movement at the beginning of words that are analysed as bearing this level tone. Words with initial obstruents tend to show a initial fall, as with the word jan³ ‘sun’, which has a pitch range of 180-150-150Hz, but the vowel is clearly level and was regarded as level by the informants. With nasal initials, there was sometimes a rise shown on the initial, as in the word nim³ nim² ‘owl’ shown in Table 7:
The tone and syllable structure of Turung

Table 7: Pitch chart for nim³ nim³ ‘owl’.

In Table 7 the slight rise at the beginning of the first syllable followed by a level coda can be clearly seen. This rise occurs at the same time as the release of the initial nasal consonant. The second syllable, which carries the prominent 2nd tone, is much more strongly released, as can be seen from the waveform in the upper portion of the table. This syllable commences with the rise at about 0.500 seconds, and reaches a peak at above 200Hz and then sharply falls.

Due to the prominence of this 2nd tone, it is usually not difficult to distinguish it from the other tones, but distinguishing the level tone (Tone 3) and the low falling tone (Tone 1) is more problematic. First of all, many words that carry what is perceptibly a level tone nevertheless show some diminution in pitch.

Secondly, although the words that are analysed as carrying the low falling tone tend to be in the lowest pitch range of an individual speaker, and those words that are analysed as being level tone are more in the mid to high range of that speaker, there are many cases where it is difficult to tell. Whereas Tai Aiton speakers, when given examples of the tones such as those presented in Table 5 can assign any new word to one or other of the tonal categories (see Morey 2005), Turung speakers generally are not able to say whether two words have the same tone or not. When Turung speakers are asked whether the words for ‘ear’ and ‘tooth’ in Table 8 have the same tone or not, their usual response is that they are “just a little bit different”. This will not necessarily mean that there is a tonal distinction. It may be vowel height, quality or other factors that influence their perception.

In Table 8, several words are presented with analysis of pitch contour done using Speech Analyzer.

---

6 The sharp rise shown at the end of the first pronunciation represents an extraneous sound.
Stephen Morey

Table 8: Comparison of the low falling tone (1st) and level tone (3rd) in Turung

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>pitch contour (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mongoose</td>
<td>jon¹</td>
<td>140-115</td>
</tr>
<tr>
<td>ear</td>
<td>naa¹</td>
<td>145-130</td>
</tr>
<tr>
<td>turtle</td>
<td>tau³</td>
<td>180-180</td>
</tr>
<tr>
<td>stomach</td>
<td>kan³</td>
<td>190-180</td>
</tr>
<tr>
<td>tooth</td>
<td>waa³</td>
<td>(120)-160-160-150</td>
</tr>
<tr>
<td>face</td>
<td>man³</td>
<td>(140)-160-160</td>
</tr>
</tbody>
</table>

The first two words are clearly falling, and clearly low. The fall is not as steep as that of words with the 2nd tone, but both in terms of pitch and contour they contrast with the second pair of words: whereas ‘mongoose’ and ‘ear’ have pitches that are in the lower part of Kon Kham’s speech range, ‘tortoise’ and ‘stomach’ are in his mid range. The contrast between these words seems clear, but what are we to make of the last two words, ‘tooth’ and ‘face’, which are somewhat lower than ‘tortoise’ and ‘stomach’ but still level? Both have voiced initials and this appears to explain the lower pitch. Since Kon Kham himself characterised the word for ‘tooth’ as a level tone, and distinct from a word with a falling contour, these words are also analysed as level.

When we examine disyllabic words neither of whose syllables carry the 2nd tone, we see considerable variation in pitch between the two syllables that it appears can only be explained by positing a distinction between a low (falling) and mid (level) tones. Some such words are listed in Table 9.

Table 9: Disyllabic words showing the comparison of the low falling tone (1st) and level tone (3rd) in Turung

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>σ1 pitch contour (Hz)</th>
<th>σ2 pitch contour (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>valley</td>
<td>ning³ gam¹</td>
<td>160-160</td>
<td>160-120</td>
</tr>
<tr>
<td>cave</td>
<td>pung³ khoo³</td>
<td>170-180-170</td>
<td>190-170-170</td>
</tr>
<tr>
<td>spring</td>
<td>khaa¹ bonq³</td>
<td>150-115</td>
<td>150-140</td>
</tr>
<tr>
<td>lion</td>
<td>hang³ sii¹</td>
<td>175-185</td>
<td>165-145</td>
</tr>
</tbody>
</table>

Tones in this paper will be marked according to the principles set out in (1) following the form recorded by Kon Kham Turung where possible.

4. Turung syllable structure

4.1 Monosyllables with a single consonant onset

In the analysis of Turung syllable structure presented here, phonetically the minimal Turung syllable is one that has an optional initial consonant, a vowel nucleus marked by tone, and a coda which may be a vowel, a nasal, or a stop consonant, as in (Error! Reference source not found.).

4) (C) V+T+ \[ V \\
S \\
N \]
In monosyllables, vowel final words are realised with longer vowels than the vowels of words that have a coda. In his initial recording of the tones (Table 5), Kon Kham Turung pronounced \( waa^3 \) ‘tooth’ with a vowel 0.39” in length and \( wa^3 \) ‘make (basket)’ with a vowel of 0’19” length. If \( waa^3 \) ‘tooth’ were pronounced with a short vowel, even without a final glottal, it would probably be misinterpreted by Turung speakers. Phonetically at least, \( waa^3 \) ‘tooth’ should be regarded as CVV, and in this paper will be written with a double vowel. Diphthongs such as in \( sai^2 \) ‘blood’ also conform to the requirements of (4). Under this analysis, we can view the Turung monosyllable as bimoraic.

In (4) the initial consonant was described as optional. For Jingpho, some scholars, such as Matisoff (1974b,2003) write a glottal stop in front of an initial vowel, where Dai (1992) and Maran (1971) omit the glottal stop and write vowel initial words. No claim is made by Matisoff and others who write initial glottal stops that there is a contrast between vowel initial words and those with glottal stops, in other words no putative contrast between \( ang \) and \( ?ang \).

One piece of evidence favouring phonemic status for initial glottal stop might be that words like \( ma\, een^2 \) ‘saliva’ are realised as [mêʔēn], in which the vowel and tone of the major syllable replicated on the minor syllable (see below Section 4.3 for a discussion of this phenomenon) and with a glottal stop. It could be argued that this glottal stop shows that the originally meaningful syllable \( een^2 \) should be analysed as having a glottal initial. This word is written by Dai as \( m\, ā^3\, jen^3 \) but as \( ma\, yēn \) by Matisoff (2003:114), the latter stating that in the speech of LaRaw Maran, the initial consonant of the major syllable is a preglottalised sonorant. As with \( peen^3 \) ‘to fly’, discussed above, this is one of the cases where Turung does not preserve an initial semivowel that is found in Jingpho, and that as a result the vowel of the major syllable is realised as a long vowel.

The glottalisation observed in the word for ‘saliva’ in Turung is much weaker than the glottal stop in \( kha^3\, sin^2 \) ‘bathe’, which is often realised phonetically as a trisyllabic structure [khâʔ ?a sîn]. Moreover, sometimes a vowel final word followed by a vowel initial word show vowel coalescence, as with \( naa^2\, ang^3 \) ‘ricefield-at’ that can be realised as [naŋ]. These two pieces of evidence would suggest that the glottal stop not be posited as an initial phoneme in Turung.

4.2 Initial clusters

In Turung, in addition to monosyllables, disyllables and other multisyllabic words, the latter of which are not discussed in detail in this paper, there are three type of words that may be regarded as longer than a syllable, but shorter than a disyllable. These are:

4) Words with initial clusters, as \( kraa^2 \) ‘hair of the head’
Sesquisyllabic words, as \( lə\, tā^2\) ‘arm’
Nasal syllables, as \( n^\prime\, lung^2 \) ‘stone’

The term sesquisyllable was first devised by Matisoff (1973:86) to refer to an iambic structure consisting of a short syllable (minor syllable) followed by a long syllable (major syllable). In Turung, most sesquisyllabic words can be realised with initial clusters, as will be discussed below. Given this, phonetic realisation alone is
therefore not enough evidence to posit that a word has a phonemic initial cluster, and we will need to establish criteria to decide whether a word is to be analysed as having an initial cluster or being sesquisyllabic. The first criterion will be the phonetic nature of the cluster. In the analysis of Turung presented here, only those clusters that have a second member that is a semivowel or a liquid will be recognised as phonemically clusters. The second criterion is that if a word is reconstructed with a cluster in proto-Tibeto-Burman, then we can recognise it as a cluster in Turung.

In the case of krea2, the proto form is given as *kra by Matisoff (2003) and *(s)-kra by Benedict (1972). Furthermore, the initial cluster has the rhotic /r/ as its second member, and thus meets both criteria to be analysed as having an initial cluster. However, in Jingpho as reported by Hanson (1906), the word is given as <kāra>, suggesting a sesquisyllabic reading.

A second example of a cluster is the word sə brang2 ‘young man’, analysed as a sesquisyllable with an initial cluster in the major syllable, an analysis supported by Hanson’s notation as <Shābrang>. If we were to analyse this word as a single syllable, it would require a triconsonantal onset, and if we were to analyse it as not having clusters at all, it would require positing two minor syllables.

This latter analysis is possible. Matisoff (2003:149) has pointed out that words with two short syllables, that is double sesquisyllables, are found in Hanson (1906); e.g. <lāsāwi> ‘bone marrow, kind of bamboo, whittle off’ and <pāsāwi> (varying with <bāswi>) ‘plaid cloth’. The variation in transcription given by Hanson for this latter word suggests that he also found the analysis of this word problematic: it being either a double sesquisyllable or a minor syllable followed by a major syllable with an initial cluster.

When Turung speakers themselves write their language in Assamese script, which has no schwa, they generally write the vowel of the sesquisyllable as the same as the vowel of the main syllable. This reflects one of the allowable pronunciations of these words (see below Section 4.3. The Assamese writing system contains many initial consonant clusters, so that we would expect cluster-initial words to be written with a cluster. However, as Table 10 shows, words which have initial clusters, such as krʊ2 ‘six’ are written as two syllables in the same way as words that are analysed as sesquisyllables. In the case of krʊ2 ‘six’, possible pronunciations vary in the same way as with sesquisyllables (see (6) below).

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Turung word</th>
<th>Written form in Assamese</th>
<th>Hanson</th>
<th>Proto form (Benedict 1972)</th>
</tr>
</thead>
<tbody>
<tr>
<td>three</td>
<td>mə sum3</td>
<td>&lt;mu sum&gt;</td>
<td>Māsum</td>
<td>*g-sum</td>
</tr>
<tr>
<td>four</td>
<td>mə lii3</td>
<td>&lt;mi li&gt;</td>
<td>Māli</td>
<td>*b-liy = *b-lay</td>
</tr>
<tr>
<td>five</td>
<td>mə nga41</td>
<td>&lt;ma nga&gt;</td>
<td>Mānga</td>
<td>*l-ŋa ~ *b-ŋa</td>
</tr>
<tr>
<td>six</td>
<td>krʊ2</td>
<td>&lt;ku ru&gt;</td>
<td>Kru</td>
<td>*d-ruk</td>
</tr>
</tbody>
</table>

4.3 Sesquisyllables

In the languages of the Jingpho group, the vowels of the minor syllable in a sesquisyllabic structure are interpreted in a variety of ways: with a schwa, as in Matisoff (1974b, 2003), followed here, or with a short <ā> symbol, following Hanson (1906),
and employed in his phonetic transcriptions by Dai and others in the *Jingpho Miwa Ga Ginsi Chum* (Jingpho-Chinese Dictionary). In the Kachin writing system now in use in for Jingpho in Kachin State of Burma, the vowel of the minor syllable is simply written as <a>.

Sometimes the minor syllable of the sesquisyllable seems to be a meaningful element, possibly reflecting an earlier stage of the language when some of these were compounds. In (5), consisting of several words relating to body parts, the minor syllable is a reduction of proto-Tibeto-Burman */l(y)ak* arm:

5)  

i)  

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>lə yung</td>
<td>‘finger’</td>
</tr>
<tr>
<td>lə saʔ</td>
<td>‘artery’</td>
</tr>
<tr>
<td>lə taʔ</td>
<td>‘arm’</td>
</tr>
</tbody>
</table>

Sesquisyllables such as those in (5) can be realised in three ways,  

i) as single syllables with initial clusters,  

ii) as sesquisyllables where the vowel of the minor syllable is a schwa, or  

iii) as sesquisyllables where the vowel of the minor syllable is in harmony with the vowel of the major syllable, or with the initial consonant of the major syllable if that is a semi-vowel.

The word for finger can thus be realised as any of the three possibilities in (6):

6)  

i)  

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ljung]</td>
<td>‘finger’</td>
</tr>
<tr>
<td>[lə jung]</td>
<td>‘finger’</td>
</tr>
<tr>
<td>[li jung]</td>
<td>‘finger’</td>
</tr>
</tbody>
</table>

The cluster realisation is particularly frequent in fast natural speech and can lead to the phonetic realisation of clusters that do not meet the first criterion for clusters given above. For example, kinship terms in Turung are all monosyllabic words that are realised with one of two prefixes:

7)  

i)  

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>‘prefix denoting one’s own relative’</td>
</tr>
<tr>
<td>k(ə)</td>
<td>‘prefix denoting another’s relative’</td>
</tr>
</tbody>
</table>

When the *k(ə)* prefix is used with one of these words, in fast speech the onset is often an initial cluster. For example, the word for ‘elder brother’, the major syllable of which is *phuu*², can be realised as *[kpʰʊː]*, with an initial cluster, even in utterance initial position. Gemination occurs when the initial consonant of the kinship term is /k/, as with *kə kuu*² ([kːʊː]) ‘someone else’s brother-in-law’. This realisation with geminate initial has only been recorded following the vowel final possessive marker *naa*³.

Although at the phonetic level the number of initial clusters in Turung is very large indeed, at an abstract level, these are regarded as the reduction of sesquisyllables, so that a large list of phonemic initial clusters is not posited for Turung.

A major question remains to be investigated: Is the minor syllable a tone-bearing unit? Writing about Jingpho, Matisoff (2003:98) said that “even though the vowels in all these minor syllables are the same, and unstressed to boot, it has been claimed (e.g. by Maran ... a native speaker) that they bear a two-way tonal contrast”. Matisoff himself
Stephen Morey

added that he had “never perceived any such contrast in Maran’s speech”, going on to point out that in Dai’s 1983 dictionary of Jingpho, three tonal contrasts were distinguished on the minor syllables.

So far in this paper, the minor syllable has not been marked for tone. When analysed by Speech Analyzer, pitch generally only shows up when the initial consonant of the minor syllable is voiced, so that the pitch that is realised may only reflect the pitch of that voiced consonant rather than the vowel which would be expected to be segment carrying contrastive tone. In the case of all the words analysed for this paper where the minor syllable was realised with a schwa, the pitch of that minor syllable was low. This matches Matisoff’s observation that in Dai’s dictionary the low tone /31/ was most frequent for the minor syllable (2003:98). If we accepted that the tones were indeed low, we might mark them all as tone 1 here. In those cases where the sesquisyllable is realised as a monosyllable with initial cluster, we would then need to suggest that the tone has been lost.

On some occasions when sesquisyllabic words are realised, the minor syllable bears a vowel that is in harmony with the vowel of the major syllable, and they do bear tone. The word mə sin³ ‘ice’ is often realised with a clear instantiation of /i/ as the vowel in the minor syllable, and that minor syllable being slightly longer than if it were realised with a schwa. This is shown in the following pitch diagram made using Speech Analyzer:

Table 11: Pitch chart for mə sin³ ‘ice’

As discussed above in Section 3, with the high level tone, (Tone 3), obstruent initials such as /s/ are often realised with a slight fall and nasal initials with a rise, as seen in this diagram. When this is taken into account, we can see that this word is realised as [mi sin], with both syllables bearing a mid level tone. It is argued here that even though the minor syllable in this example is clearly tone bearing and fully vocalised, the vocality has come by harmony with the major syllable, and the tone by
The tone and syllable structure of Turung

spreading from that syllable. In other words both vowel and tone in the minor syllable match those in the major syllable. No contrastive tone for the minor syllable is posited.

Some disyllabic words are also pronounced with short initial syllables, and are thus quite similar to some of the realisations of sesquisyllables. For example, pu³ sai² ‘freshwater eel’, was pronounced in elicitation with a short initial syllable (0’12”) and a much longer second syllable (0’41”). The first syllable had a hint of final glottal closure and the structure was very similar to a sesquisyllable. However, unlike the pronunciation of mo sin³ ‘ice’ discussed above, neither the vowel nor the tone of the short syllable reflected the vowel of the major syllable. This word, then, is analysed as a multisyllabic word, made up of two full syllables. It may, over time, come to be regarded as sesquisyllabic, in which case we would expect pronunciations like [pə səi] or [pə səi].

4.4 Nasal Syllables

The third type of syllable presented above in (4) is here called a ‘nasal syllable’, consisting of a minor syllable that is a syllabic nasal, followed by a major syllable, as in n¹ lung² ‘stone’. In all the examples recorded in Turung so far, the initial consonant of the major syllable may be a stop, fricative or semivowel, but not a nasal consonant, unlike in Jingpho (see Table 13 below). Phonetically, the nasal assimilates to the place of articulation of the major syllable, and will here be written in this assimilated form, following the expressed wishes of the Turung community.

Table 12 presents examples of this type of syllable. In each case the minor syllable, the syllabic nasal, is analysed as having the 1st tone, the low falling tone, although the actual profile of the tone varies slightly, and to some extent reflects the profile of the tone of the main syllable.

Table 12: Turung nasal syllables.

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>pitch contour of syllabic nasal</th>
<th>pitch contour of major syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>ground</td>
<td>ng¹ gaa¹</td>
<td>160-150</td>
<td>160-140</td>
</tr>
<tr>
<td>stone</td>
<td>n¹ lung²</td>
<td>150-160</td>
<td>150-200-120</td>
</tr>
<tr>
<td>air</td>
<td>m¹ bung³</td>
<td>150-160</td>
<td>170-160</td>
</tr>
</tbody>
</table>

In the case of ng¹ gaa¹ ‘ground’, the nasal syllable falls slightly, in line with the main syllable; whereas n¹ lung² ‘stone’ shows a slight rise is perceptible, suggesting a tendency to reflect the rising-falling profile of the major syllable.

Matisoff (2003:129) discussed nasal syllables in Jingpho, describing the language as “particularly interesting in this regard” and pointing out that the syllabic nasals “can bear tone”. Some examples of such words are given in Table 13, where the syllabic nasal is in variation with full syllable (“dimidiated to a full syllable, nif- or num-²”).

---

7 Matisoff (2003:129 fn103) cast doubt on whether a form like this was found in Jinghpo; it is certainly found in Turung.

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>pitch contour of major syllable (Hz)</th>
<th>pitch contour of syllabic nasal (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lizard</td>
<td>ṅ-sāŋ-sōn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edge</td>
<td>ṅ-gām / nīŋ-gām</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stone</td>
<td>ṅ-lūŋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tail</td>
<td>ṅ-mài</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axe</td>
<td>ṅ-wā / nīŋ-wā</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wound, scar</td>
<td>ṅ-mā / nūm-mā</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In most of these examples, the tone of the syllabic nasal is the same as the tone of the major syllable, the exception being the word for ‘axe’. In Turung, ‘axe’ is realised as nung¹ waa¹, with a full first syllable. No form n¹ waa¹ has been recorded. Several other Jingpho words that have been recorded as nasal syllables are only recorded with full first syllables in Turung, as for example Jingpho ṅ-sā ‘old’ which in Turung is nīng¹ saa². Interestingly, in all words where Turung has a full syllable and Jingpho has only a syllabic nasal for the first syllable, Turung has a low tone on the first syllable.

As discussed earlier, Turung does not seem to have nasal syllables where the major syllable has a nasal initial. Whereas Jingpho has ṅ-mài for ‘tail’, in Turung this word is mai³ can³, without any evidence of a syllabic nasal.

In fact, monomorphemic words with syllable nasal initials in Turung seem to fall into two groups. Several very common function words were pronounced with a high tone on the syllabic nasal, as can be seen with the first two examples in Table 14. This has tentatively been marked as Tone 2.

Table 14: Tones on Turung nasal syllables.

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>pitch contour of major syllable (Hz)</th>
<th>pitch contour of syllabic nasal (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>n² dai³</td>
<td>200-200</td>
<td>200-190</td>
</tr>
<tr>
<td>here</td>
<td>ng² goi¹</td>
<td>190-200-190</td>
<td>190-170</td>
</tr>
<tr>
<td>bow</td>
<td>n¹ dan³</td>
<td>140-150-140</td>
<td>160-160</td>
</tr>
<tr>
<td>cloth</td>
<td>m¹ baa¹</td>
<td>130-130</td>
<td>140-120</td>
</tr>
<tr>
<td>court, yard</td>
<td>m¹ pan¹</td>
<td>130-130</td>
<td>140-130</td>
</tr>
<tr>
<td>room</td>
<td>n¹ dun³</td>
<td>150-150</td>
<td>170-160</td>
</tr>
<tr>
<td>seed</td>
<td>n¹ lii³</td>
<td>160-160</td>
<td>170-170</td>
</tr>
<tr>
<td>bran, husk</td>
<td>n¹ san²</td>
<td>150-150</td>
<td>200-150</td>
</tr>
<tr>
<td>husked rice</td>
<td>ng¹ guu¹</td>
<td>140-140-120</td>
<td>150-170-160</td>
</tr>
</tbody>
</table>

The content words in Table 14 all exhibit syllabic nasals with lower tones than those found on the two function words, but are these tones all the same? In the case of the words for ‘bow’, ‘cloth’ and ‘court’, the tone of the syllabic nasal was clearly low and in the case of ‘cloth’ and ‘court’ virtually matched the tone of the major syllable and so these were listed as tone 1. Where the major syllable is tone 3, the higher, level tone, as with ‘room’ and ‘seed’, the tone of the syllabic nasal is also higher, as in the case of n¹ dun³ ‘room’. Here, however, the syllabic nasal bore a perceptually lower tone than the major syllable, and is therefore analysed as the 1st tone.
Let us turn back for a moment to the word m\textsuperscript{1} bung\textsuperscript{3} ‘air’ seen above in Table 12. The presence of a syllabic nasal here is interesting because many words relating to the weather are sesquisyllabic, with a prefix m\textcircled{a}-, such as m\textcircled{a} sin\textsuperscript{3} ‘ice’, discussed in Section 4.3 above. This prefix is arguably a reduction of the word m\textcircled{u} sin\textsuperscript{2} ‘sky’. Given that, we might expect that the word for ‘air’ also included this prefix, which has now become a syllabic nasal, rather than remaining as the minor syllable of a sesquisyllabic structure. Hanson does record this word as <Nbung>, showing that it is a nasal syllable.

There is another type of nasal syllable, the negative marker n\textsuperscript{3}. For Jingpho, Matisoff (2003:40) regarded this as a nasal prefix, although in the writing system used by Jingphos in Kachin state it is written as a separate word. When I first heard this morpheme, in the speech of an Aiton-Turung bilingual who was trying to teach me the Turung language prior to my first fieldwork in Rengmai village in 2003, I noticed that with verbs that bore the 2\textsuperscript{nd} tone (rising and high falling), the verb seemed to have a lower tone in the negative form than when it was in the positive. Unfortunately this was not recorded, and speakers do not always produce a similar effect.

Interestingly, for Jingpho, Matisoff has noted that when a verb has a low tone, the negative prefix causes it to assume the high-falling sandhi tone, as l\textsuperscript{u} ‘have’, â-l\textsuperscript{u} ‘not have’. In Turung, this verb lu\textsuperscript{2} already has a high-falling tone in citation, and as hinted earlier, there is some possibility that in Turung there is a reverse process in operation.

To check any potential effects of the negative marker on the tone of the verb, Table 15 lists three verbs with the negative form and compares the pitch contour of the verb both with and without the negative morpheme preceding it.

<table>
<thead>
<tr>
<th>gloss</th>
<th>example</th>
<th>pitch contour of negative nasal (Hz)</th>
<th>pitch contour of verb (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cooked</td>
<td>s\textsuperscript{a} duu\textsuperscript{3}</td>
<td>σ\textsubscript{1} 130-120; σ\textsubscript{2} 160-150</td>
<td></td>
</tr>
<tr>
<td>uncooked</td>
<td>n\textsuperscript{3} s\textsuperscript{a} duu\textsuperscript{3}</td>
<td>160-140</td>
<td>σ\textsubscript{1} 130-130; σ\textsubscript{2} 170-160</td>
</tr>
<tr>
<td>ripe</td>
<td>min\textsuperscript{3}</td>
<td>(130)-170-170</td>
<td>170-170</td>
</tr>
<tr>
<td>unripe</td>
<td>n\textsuperscript{3} min\textsuperscript{3}</td>
<td>170-170</td>
<td>170-170</td>
</tr>
<tr>
<td>good taste</td>
<td>muu\textsuperscript{3}</td>
<td>(120)-170-170</td>
<td></td>
</tr>
<tr>
<td>tasteless</td>
<td>n\textsuperscript{3} muu\textsuperscript{3}</td>
<td>140-160</td>
<td>170-170</td>
</tr>
</tbody>
</table>

As can be seen, there is no appreciable effect on the tone of the verb from the negative nasal. It has been assigned the level tone here, in particular in view of the form n\textsuperscript{3} min\textsuperscript{3} ‘unripe’ where the tone on the negative nasal and the verb are identical. These two morphemes were realised with continuous articulation, but a clear pulse between the two nasals segments (that of the negative marker and the initial consonant of the verb) is clearly audible.

On the other hand, n\textsuperscript{3} muu\textsuperscript{3} ‘bland’ was realised with a long, geminate [m] as the initial consonant, as [muu]. It is only the length that signalled the negation\textsuperscript{8}. The length of the initial nasals was measured, using Speech Analyzer, for this word in both negative and positive, as shown in (8):

---

\textsuperscript{8} There are a number of cases in the Turung texts where only length of the initial nasal segment signals the negative. This has sometimes led to misunderstanding by the present writer.
Stephen Morey

8) muu³ ‘good taste’ [mu:] 0’14”
   n³ muu³ ‘tasteless’ [mu:] 0.19”

The negative initial nasal is significant longer, nearly one and a half times the length of that in the positive verb.

One further consideration is the possibility that a sesquisyllable with a minor syllable mₘ might come to be realised as a nasal syllable. When Kon Kham pronounced the word for ‘bruise’, phiʔ² mₘ luʔ¹ quickly, it was phonetically realised as [phi m lû], suggesting a syllabic nasal. However, this word is analysed as being a sesquisyllable, firstly because another informant pronounced the word clearly as [phi m lû], and secondly because if it were a syllabic nasal, the place of articulation of the nasal would have assimilated to the initial consonant of the final syllable, being realised as [phi n lû].

5. Notes on stress in Turung

One difficulty in the tonal analysis of Turung is that tones are affected by their position in a phrase or utterance. Compare the tokens of the word ding³ laa¹ ‘old man’ in three contexts, compared in Table 16.

| Table 16: Effect of neighbouring segments on the tone of ding³ laa¹ ‘old man’ |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|
| example | gloss | pitch contour of 1st syllable | pitch contour of last syllable | pitch contour of waa¹ ‘DEF’ | pitch contour of waa² ‘return’ |
|         |      | of ding³ laa¹ ‘old man’ (Hz) | of ding³ laa¹ ‘old man’ (Hz) | (Hz) | (Hz) |
| ding³ laa¹ | ‘old man’ | 150-150 | 150-135 | - | - |
| ding³ laa¹ waa¹ | ‘the old man’ | 140-130 | 140-160-140 | 140-110 | - |
| ding³ laa¹ waa¹ waa² | ‘the old man went back’ | 140-140 | 140-140 | 140-140 | 140-160-100 |

The recording of this word in citation was part of the recording of the longer word list, whereas the two examples of the word in context were from the original analysis of the tones presented above in Table 5, where Kon Kham Turung’s pitch range was generally lower than it was in the word list.

Taking this into account, we suggest that the first syllable of ding³ laa¹ does not change significantly from one example to another, but that the second syllable does. When followed by the definiteness marker, a syllable which also carries the low falling, 1st, tone, the last syllable of ding³ laa¹ shows higher pitch and falling contour, perceptually similar to the 2nd tone. In the last example, on the other hand, the syllables were all level and any contrast between the 1st and 3rd tone was neutralised.

The word sum¹ phoʔ¹ ‘person’ was pronounced in elicitation with two very similar tones, both falling from about 160 to 140 Hz. However, when Kon Kham combined this word with the definiteness marker waa¹, the second syllable was significantly higher than the first, and was perceptually similar to the 2nd tone.
The tone and syllable structure of Turung

It is argued here that when a noun is followed by the definiteness marker, it forms a single phonological word, especially when only the noun and definiteness marker are present. Such phonological words behave like compounds. In compounding of more than two syllables, two processes occur. The first is a natural tendency towards sesquisyllabic structures, where the first syllable is weakened. The second is adding stress to the penultimate major syllable.

For example, in elicitation, _mung^2_ ‘country’ has a clear rising-high-falling profile, with a pitch contour of 120-200-100 Hz. However, in the compound _mung^2_ _kang^2_ _gaa^1_ ‘earth, land’, the first syllable (150-170-160Hz) is much lower than the second (190-170Hz) a syllable that is also pronounced much more strongly. We would argue that this can be explained, as can the examples with the definiteness marker above, as cases of penultimate stress.

This penultimate stress applies to the penultimate major syllable. The word _lo gong^3_ _lo yung^3_ ‘toe’ is made up of two sesquisyllables, both of which in elicitation consisted of an unstressed first syllable and a second syllable that was level at a pitch of about 180Hz. In the compounded form, however, the second syllable, _gong^3_ was clearly higher in tone and more stressed than the last syllable _yung^3_.

The tendency for disyllabic words to reduce to sesquisyllabic structures means that sometimes the anti-penultimate syllable is stressed, as with _ng^1_ _gaa^1_ _sum^3_ _mun^1_ ‘dust’, where the stressed syllable is _gaa^1_ which was realised as a rising tone, reaching up into the higher levels of Kon Kham’s pitch range. In citation of the word _ng^1_ _gaa^1_ ‘ground’, the tone is low falling. This analysis is supported by the fact that in at least some of Kon Kham’s pronunciations of the word, the syllable _sum^3_ was reduced to [sa].

6. Turung tones – a reprise

In Section 3 above, we posited three tones for monosyllables in Turung, listed in (1) above. In Section 4.4 we further suggested that there are three tones for syllabic nasal syllables, but that the profile of these tones was different from the tones of monosyllables. The two systems are compared in Table 17:

<table>
<thead>
<tr>
<th>Tone No.</th>
<th>Example</th>
<th>Gloss</th>
<th>Tone contour (Hz)</th>
<th>Example</th>
<th>Gloss</th>
<th>Tone contour of σ1 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>jon^1</td>
<td>mongoose</td>
<td>150-120</td>
<td>m^1 baa^1</td>
<td>cloth</td>
<td>130-130</td>
</tr>
<tr>
<td>2</td>
<td>sai^2</td>
<td>blood</td>
<td>200-120</td>
<td>n^2 dai^3</td>
<td>this</td>
<td>200-200</td>
</tr>
<tr>
<td>3</td>
<td>tau^3</td>
<td>turtle</td>
<td>180-180</td>
<td>n^3</td>
<td>NEG</td>
<td>170-170</td>
</tr>
</tbody>
</table>

The syllabic nasals consist of three level tones, the 1st tone being low, the 2nd tone high and the 3rd tone mid. The tones on monosyllables differ, in that both the low and high tone are falling tones, whose beginning pitch is similar to that of the corresponding syllabic nasal.

We might then ask where the contours for monosyllables have come from? The three contours listed in Table 17 are very similar to the three tones of Aiton, with which Turung has long been in contact (see Morey 2005:160f) and perhaps the realisation of tone on monosyllabic words has been influenced by Aiton. In view of this, it is worth
noting that Turung speakers admit that ‘the tune’ is one of the salient differences between their variety and that of the Singphos of Upper Assam.

And what of parallels between our analysis of Turung tones and the tonal systems of the neighbouring Jingpho language? There does appear to be one parallel: a high proportion, though by no means all, of the words that Matisoff (1974b) lists as low tone for Jingpho are found with the 2nd tone (high) in Turung.

The Jingpho mid tone, on the other hand, corresponds with all three Turung tones, although it is most commonly found corresponding with the 3rd tone and 1st tones, as shown in Table 18.

**Table 18: Comparison of Jingpho mid tone in Matisoff (1974b) and Turung.**

<table>
<thead>
<tr>
<th>Matisoff No.</th>
<th>English gloss</th>
<th>Jingpho Mid Tone</th>
<th>Turung</th>
<th>pitch contour (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>‘fly (v.)’</td>
<td>pyĕn</td>
<td>peen³</td>
<td>170-170</td>
</tr>
<tr>
<td>6</td>
<td>‘moon, month’</td>
<td>tā, sōtā</td>
<td>sō taa¹</td>
<td>160-135</td>
</tr>
<tr>
<td>12</td>
<td>‘smoothed out’</td>
<td>prî</td>
<td>a¹ prii³</td>
<td>180-210-100</td>
</tr>
<tr>
<td>20</td>
<td>‘ant’</td>
<td>ūkîyîn</td>
<td>gu³ gin³</td>
<td>190-190</td>
</tr>
<tr>
<td>26</td>
<td>‘prison’</td>
<td>thõŋ</td>
<td>thong²</td>
<td>185-200-110</td>
</tr>
<tr>
<td>76</td>
<td>‘barking deer’</td>
<td>khyî</td>
<td>cə khii³</td>
<td>200-180-180</td>
</tr>
<tr>
<td>85</td>
<td>‘ten’</td>
<td>sî</td>
<td>sii³</td>
<td>180-180</td>
</tr>
<tr>
<td>101</td>
<td>‘die’</td>
<td>sî</td>
<td>sii³</td>
<td>180-180</td>
</tr>
<tr>
<td>111</td>
<td>‘dream’</td>
<td>māŋ</td>
<td>(yup²) mang¹</td>
<td>150-120-120</td>
</tr>
<tr>
<td>134</td>
<td>‘hundred’</td>
<td>lōtsā</td>
<td>la caa¹</td>
<td>160-140</td>
</tr>
</tbody>
</table>

The problem of the tonal analysis of Turung is exacerbated by the fact that tonal inconsistencies among the Turung are not uncommon. When eliciting the word for ‘vomit’ *m¹ phat¹*, the major syllable was pronounced in two different ways, once as low falling, and once as a much higher falling tone. One minimal pair which some old Turung speakers pronounced is given in (9)

9) sat¹ ‘rice’
   sat² ‘kill’

When asked about this contrast, Kon Kham expressed the view that the two words were the same, and when talking about them with other Turung informants, referred to the first alternative as in (10):

10) saa² dii³ sat¹ phee³ ai¹ khii³ san²
   eat GV rice A.AG PRT 3SG ask

   ngaa² waa¹ HAVE DEF

   ‘It is the sat¹ which we eat that he is asking about.’

The phrase *saa² dii³ sat¹* ‘the sat¹ which we eat’, is used to distinguish this word from the other *sat²* ‘to kill’. If Kon Kham regarded these two words as having contrastive tones, we would not expect him to need to refer to *sat¹* ‘rice’ in this way. So
The tone and syllable structure of Turung

it is that Turung speakers do not all agree on the tones of particular words. Readers of this paper will also have gathered that modern researchers on languages of the Jingpho group do not always agree on the analysis either. It might, in view of this, be appropriate to end with a quote from Ola Hanson, who lived many years with Jingpho speakers:

“The difficult problem regarding the tones may seem to have been entirely ignored, as no tonal marks have been introduced. ... The tones are more important than generally admitted by Kachin students, but they can be mastered only with the help of a native teacher, and it would be useless to burden these pages with tonal marks in regard to which no two Europeans would ever agree.” (Hanson 1906:iv bold mine)

7. Bibliography


THE VIETO-KATUIC HYPOTHESIS: LEXICAL EVIDENCE

Mark J. Alves
Montgomery College
markalves2004@yahoo.com

1. Introduction
The position of linguistic affiliation of Vietnamese within Austroasiatic has been proven satisfactory to most researchers.\(^1\) Lexical evidence (Huffman 1977, Nguyễn T.G. 1978) and phonological data (Haudricourt 1954, Ferlus 1976 and 1981, Diffloth 1989, Nguyễn T.C. 1995) have, using standard historical linguistic reconstructive methodology, solidly demonstrated a genetic linguistic connection between Vietnamese and the Mon-Khmer branch of Austroasiatic.

It is in particular the evidence seen in more conservative varieties of Vedic, including the Minor Vedic/P’ọng-Chứt languages,\(^2\) varieties of Mường, and even North Central Vietnamese (Nguyen T.C. 1995, Alves (forthcoming)), that links the heavily sinified speech of traditional northern Vietnamese with Mon-Khmer. In addition, scholars have begun to suggest in print the close link between Vedic and Katuic in particular (Diffloth 1989, 1990, and 2004), and indeed, a proto-Vieto-Katuic grouping has been posited (Diffloth 1990, Nguyễn T.C. (1995)). On the place of Vietnamese within Mon-Khmer, Diffloth (ibid.: 126) wrote the following.

“Haudricourt was elusive on the matter. But Ferlus (1991) suggested “des solides affinités” between Katuic, Bahnaric, and Việt-Mường without providing any evidence. He was presumably thinking about the lexicon, and I would share that impression, especially with regards to Viet and Katuic.”

The focus of this paper is precisely the lexical evidence (see Sections 4 and 5) that links the Vedic and Katuic branches of Mon-Khmer, with the resulting discovery of possible additional etymological layers between Austroasiatic and Vedic. The supporting lexical data is discussed in more detail below. First, however, in Sections 2 and 3, the data sources and technique for analysis are discussed and a few of the phonological correspondences on which those lexical forms rely for confirmation are summarized.

\(^1\) However, various researchers have different views on the matter, suggesting origins in a kind of Austro-Thai grouping and including Japanese (Matsumoto 1928 and Nguyễn N.B. 1994), a minor Mon-Khmer substratum with Tai as the main source of the lexical and structural elements of Vietnamese (Maspero 1912, and Hoàng T.C., personal communication), or a language without precise genetic linguistic affiliation (which appears to be the official stance in Vietnam since lists of ethnic groups grouped by linguistic families, but Vietnamese is listed on its own, separate from other Mon-Khmer languages).

\(^2\) There is no current standard usage to refer to the group of Vedic languages outside of the Việt-Mường subbranch. Alves (2003), in a SEALS VII presentation in 1997, used the term ‘Minor Vedic’, while Nguyễn T.C. (1995) uses the term ‘P’ọng-Chứt’. The latter term is useful since it is used to distinguish that group for historical grouping and still captures roughly the geographic division.
2. Data and Analysis

The data for this study consist of about 280 fairly solid cognates within Austroasiatic or some sub-branch therein. Of these 280, about 40 are categorized as strong candidates for Vieto-Katuic cognates (with a few dozen more of less certain status). These 40 forms show common phonological patterns of changes and innovations and are generally restricted to Vietic and Katuic (with some minor exceptions in neighboring languages and possible loanwords), as discussed below.


The lexical data were analyzed, and etymological layers based on proto-levels were identified, including (a) Austroasiatic, (b) Mon-Khmer, and (c) possible subgroups within Mon-Khmer (Southern, Northern, and Eastern). Triangulation of the appearance of words (i.e., three sub-branches or three languages within a sub-branch) was considered the minimum to classify lexemes as solid cognates. Without triangulation, the forms were not considered viable candidates until such time that other data are discovered. Within Vietic, the languages Vietnamese, Ruc, and Thavung were the primary sources. These were compared with a Proto-Katuic reconstruction (Sidwell ibid.), but in some cases, words that were not reconstructed by Sidwell but which at least three Katuic languages had potential cognates were considered viable. In general, the point was to provide a relatively conservative method of excluding data.

A few dozen other items were either excluded or treated differently due to confounding factors. These included words that appear onomatopoetic (e.g., ‘cut’, ‘hit’, and ‘sip/suck’) or those that could be loans from Tai-Kadai or Sinitic (Pou and Jenner 1973 was the primary reference) and words seen among language families in Southeast Asia (such as ‘this’ and ‘eye’) are also excluded or kept on the list as weaker evidence. Such words are kept but not used in strong statements about genetic linguistic affiliation.

Another complication is that certain posited Vieto-Katuic forms could be borrowings or look-alikes. However, with more than forty forms, it is assumed here retention is the most likely explanation in the majority of those situations. In the end, while certain posited forms may turn out not to be proto-Vieto-Katuic cognates, and in fact, a proto-Vieto-Katuic phonological system has yet to be generated, the connection between Vietic and Katuic still seems a reasonable hypothesis for the time being.

The data also suggest two other possible subgroups: (a) a larger Northern and Eastern group and (b) a Bahnaro-Vieto-Katuic (BVKa) subgroup within Eastern Mon-Khmer. If the data are proven viable, the BVKa group would appear to some kind of non-Monic and non-Khm eric group that developed in the modern region of Northeast Thailand, Southern Laos and Central Vietnam. Certainly, this hypothesis would require additional support from phonological reconstructions as well as some idea of the contact situation.
3. Phonological Support

The phonological evidence for a Vieto-Katuic subgroup within Eastern Mon-Khmer includes five main points. There is a combination of evidence seen among consonants, phonation, and syllable structure. For each aspect, specific lexical items are taken from the dataset table in the next section, and the numbers of the lexical items refer to their location in the table. The five issues are as follows.

a) Proto-Mon-Khmer */ʔ/ realized as Proto-Vieto-Katuic */s/ and */h/ (Diffloth 1990) (e.g., #2 ‘blood’ and #3 ‘bone’)

b) The preservation of Proto-Mon-Khmer */c/ in both Vietic and Katuic (in contrast with the innovative Proto-Bahnaric and Northern Khmer */s/) (reinterpretation of data presented in Ferlus 1978)³ (e.g., #1 ‘bird’, #19 ‘dog’, and #43 ‘ripe’)

c) Creaky voice in Katuic and in Vietic with open syllables having sác/nāng tones (Diffloth 1989) (e.g., #9 ‘one’, #22 ‘four’, and #35 ‘bitter’)

d) Adding presyllables to proto-Mon-Khmer forms (Nguyễn T.C. 1995) (e.g., #5 ‘fish’, #24 ‘hair’, and #25 ‘leaf’)

4. Etymological Layers

The following pages in Section 5 contain 100 sample words from the various posited etymological levels from Austroasiatic to Vieto-Katuic.⁴ The layers and the number of the lexical items in the following chart are shown below.

(a) AA = Austroasiatic (#1-13)
(b) MK = Mon-Khmer (#14-33)
(c) NEMK = Northern and Eastern Mon-Khmer (#34-46)
(d) EMK = Eastern Mon-Khmer (#47-59)
(e) BVKa = Bahnaro-Vieto-Katuic (#60-68)
(f) VKa = Vieto-Katuic (#69-100)

Supplementary support for the division of etymological layers comes from examples of semantic specialization. Higher level items in these examples are generic terms, whereas the subbranches have additional semantically specialized terms.

1. Mon-Khmer ‘tooth’ in #31, Eastern Mon-Khmer ‘canine teeth’ in #59, and Bahnaro-Vieto-Katuic ‘gums’ in #61
2. Austroasiatic ‘arm/hand’ in #7 and Vieto-Katuic ‘armpit’ in #69.
3. Austroasiatic ‘fly (insect)’ in #6 and Vieto-Katuic ‘bluebottle fly’ in #81.
4. Austroasiatic ‘bird’ in #1, Northeastern Mon-Khmer ‘crow’ in #18, and Vieto-Katuic ‘duck’ and #77.

³ Ferlus, who attempted to show how Vietnamese phonemes patterned with Northern Mon-Khmer, in fact provided reasonable evidence for Vieto-Katuic, in particular, the shared Proto-Mon-Khmer palatal stop */c/ in both Vietic and Katuic, in contrast with the Bahnaric innovation */s/.

⁴ A complete with over 300 forms (including forms of less certain status) list is available at www.geocities.com/malves98. In addition, the complete table includes data from the Vietic Language Thavung and Proto-Bahnaric.
5. Dataset

The table below contains (1) the English gloss, (2) the etymological source level, (3) the Vietnamese form, (4) the form from the highly conservative Vietic language, Ruc, and (5) the proto-Katuic form (from Sidwell Ibid.) or forms from multiple Katuic languages when no reconstruction exists. The Ruc forms come from two sources, which are separated by a backslash. Multiple forms from the same source are separated by commas. Double ‘x’ indicates that the form was not found in available sources. Under the column with proto-Katuic, (P) stands for Pacoh, (T) for Taoih, and (B) for Bru. These are forms that have no proto-Katuic reconstruction, but for which there is still support within the language group.

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Level</th>
<th>Viet</th>
<th>Ruc</th>
<th>Proto-Kat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. bird</td>
<td>AA</td>
<td>cim 1</td>
<td>?i.cim 1</td>
<td>*ceem</td>
</tr>
<tr>
<td>3. bone</td>
<td>AA</td>
<td>saiq1</td>
<td>saq1/ saq2</td>
<td>*thaap</td>
</tr>
<tr>
<td>4. ear</td>
<td>AA</td>
<td>taj 1</td>
<td>saq1</td>
<td>*ktoor</td>
</tr>
<tr>
<td>5. fish</td>
<td>AA</td>
<td>ka3</td>
<td>?a.ka3</td>
<td>*taaa</td>
</tr>
<tr>
<td>6. fly (v)</td>
<td>AA</td>
<td>baj1</td>
<td>paar1/ paar1</td>
<td>*par</td>
</tr>
<tr>
<td>7. hand/arm</td>
<td>AA</td>
<td>taj 1</td>
<td></td>
<td>*taii</td>
</tr>
<tr>
<td>8. nose</td>
<td>AA</td>
<td>muj6</td>
<td>mujh2/ mujh2/</td>
<td>*moh</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>muh1, mulj1</td>
<td></td>
</tr>
<tr>
<td>9. one</td>
<td>AA</td>
<td>mot 4</td>
<td>moe4</td>
<td>*mooj</td>
</tr>
<tr>
<td>10. rain</td>
<td>AA</td>
<td>mis1</td>
<td>xx</td>
<td>*moo</td>
</tr>
<tr>
<td>11. root</td>
<td>AA</td>
<td>re6</td>
<td>reah2/ lierh1</td>
<td>*riaas</td>
</tr>
<tr>
<td>12. two</td>
<td>AA</td>
<td>haj 1</td>
<td>haj1</td>
<td>*baar</td>
</tr>
<tr>
<td>13. water</td>
<td>AA</td>
<td>niak3</td>
<td>daak3</td>
<td>*daak-*daak</td>
</tr>
<tr>
<td>14. 2nd pers.</td>
<td>MK</td>
<td>maj 2</td>
<td>?a.mi1, mi1</td>
<td>*maj</td>
</tr>
<tr>
<td>Sing., you</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. breathe</td>
<td>MK</td>
<td>njij5</td>
<td>ta.νas1</td>
<td>*tnjih</td>
</tr>
<tr>
<td>16. centipede</td>
<td>MK</td>
<td>ret1</td>
<td>ka.siip3/ ka.sit3</td>
<td>*kaheep</td>
</tr>
<tr>
<td>17. child</td>
<td>MK</td>
<td>kon1</td>
<td>kon1</td>
<td>*kaon</td>
</tr>
<tr>
<td>18. crow (n.)</td>
<td>MK</td>
<td>?ak3</td>
<td>xx</td>
<td>*[k?]aak</td>
</tr>
<tr>
<td>19. dog</td>
<td>MK</td>
<td>ca3</td>
<td>?a.ca3</td>
<td>*caao</td>
</tr>
<tr>
<td>20. fly (n.)</td>
<td>MK</td>
<td>ruaj5</td>
<td>ma.roaj2, pa.roaj2/ mu.raj1</td>
<td>*araoj</td>
</tr>
<tr>
<td>21. foot, leg</td>
<td>MK</td>
<td>can1</td>
<td>ciη2</td>
<td>*fian</td>
</tr>
<tr>
<td>22. four</td>
<td>MK</td>
<td>bon3</td>
<td>pon3</td>
<td>*pan</td>
</tr>
<tr>
<td>23. fruit</td>
<td>MK</td>
<td>taj3</td>
<td>pu.li3/ po.li3</td>
<td>*palaj</td>
</tr>
<tr>
<td>24. hair</td>
<td>MK</td>
<td>tauk3</td>
<td>?u.suk3</td>
<td>*sok</td>
</tr>
<tr>
<td>25. leaf</td>
<td>MK</td>
<td>la3</td>
<td>?u.la3, hla3</td>
<td>*salaa</td>
</tr>
<tr>
<td>26. house, head</td>
<td>MK</td>
<td>caj3</td>
<td>ci3</td>
<td>*pcaj</td>
</tr>
<tr>
<td>27. meat, flesh</td>
<td>MK</td>
<td>ti3it1</td>
<td>siit1</td>
<td>*sac</td>
</tr>
<tr>
<td>28. mosquito</td>
<td>MK</td>
<td>muaj6</td>
<td>kep3</td>
<td>*moos</td>
</tr>
<tr>
<td>29. new</td>
<td>MK</td>
<td>maj3</td>
<td>baj1/ baj3</td>
<td>*tmee</td>
</tr>
<tr>
<td>30. shoot</td>
<td>MK</td>
<td>ban3</td>
<td>pin3</td>
<td>*pan</td>
</tr>
<tr>
<td>31. tooth</td>
<td>MK</td>
<td>ran1</td>
<td>ka.saq1</td>
<td>xx</td>
</tr>
<tr>
<td>32. weave</td>
<td>MK</td>
<td>dam1</td>
<td>taj1</td>
<td>*taan</td>
</tr>
<tr>
<td>33. weep, cry</td>
<td>MK</td>
<td>xauk3</td>
<td>jaam4</td>
<td>*naam-*jiim</td>
</tr>
<tr>
<td>34. betel leaf</td>
<td>NEMK</td>
<td>taw2</td>
<td>plu2</td>
<td>*balua</td>
</tr>
<tr>
<td>Entry</td>
<td>Language</td>
<td>Meaning</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>bitter</td>
<td>NEMK</td>
<td>dañ⁴</td>
<td>tan⁴ / xx</td>
<td>*?atañ</td>
</tr>
<tr>
<td>deep</td>
<td>NEMK</td>
<td>šaw¹</td>
<td>ca.ru¹, jorù¹, tru¹, tu¹ / xx</td>
<td>*truul[?]</td>
</tr>
<tr>
<td>far</td>
<td>NEMK</td>
<td>sa¹,</td>
<td>ca.ña⁴j / ca.ña⁴j</td>
<td>*eñaaj</td>
</tr>
<tr>
<td>mortar</td>
<td>NEMK</td>
<td>koj³</td>
<td>to.kod³</td>
<td>*tpal</td>
</tr>
<tr>
<td>pestle</td>
<td>NEMK</td>
<td>caj²</td>
<td>?n.ri²</td>
<td>*?n⁴ree</td>
</tr>
<tr>
<td>rat</td>
<td>NEMK</td>
<td>cuat</td>
<td>ko.ne³ / ko.ne³</td>
<td>*knee</td>
</tr>
<tr>
<td>rice chaff / husk</td>
<td>NEMK</td>
<td>kaam³ (bran)</td>
<td>ki.tik³ / ka.tik³</td>
<td>*?kaam</td>
</tr>
<tr>
<td>rice, unhusked</td>
<td>NEMK</td>
<td>yaw⁴</td>
<td>ro.ko³</td>
<td>xx</td>
</tr>
<tr>
<td>ripe, cooked</td>
<td>NEMK</td>
<td>cin³</td>
<td>chin³</td>
<td>*ceen</td>
</tr>
<tr>
<td>snake</td>
<td>NEMK</td>
<td>rañ³</td>
<td>pu.sign³ / pa.sign³</td>
<td>*kañ</td>
</tr>
<tr>
<td>thatch</td>
<td>NEMK</td>
<td>tañ¹</td>
<td>?m.len¹ / mo.len²</td>
<td>*plañ</td>
</tr>
<tr>
<td>thigh</td>
<td>NEMK</td>
<td>duj²</td>
<td>pu.lu¹</td>
<td>*balaw</td>
</tr>
<tr>
<td>bark, shell</td>
<td>EMK</td>
<td>vs¹</td>
<td>ka.duk¹ /</td>
<td>*ndsh-</td>
</tr>
<tr>
<td>die</td>
<td>EMK</td>
<td>cet¹</td>
<td>ki.cit¹ / ku.cit¹, ka.cit¹</td>
<td>*kceet</td>
</tr>
<tr>
<td>grandchild</td>
<td>EMK</td>
<td>caw³</td>
<td>cu³</td>
<td>*acaw</td>
</tr>
<tr>
<td>hundred</td>
<td>EMK</td>
<td>ūtam¹</td>
<td>klam¹</td>
<td>*kalam</td>
</tr>
<tr>
<td>inside</td>
<td>EMK</td>
<td>ūtañ¹</td>
<td>klañ¹</td>
<td>*kaluñ</td>
</tr>
<tr>
<td>kill</td>
<td>EMK</td>
<td>ziñ³</td>
<td>ka.cit³</td>
<td>*kneet</td>
</tr>
<tr>
<td>lime (for betel)</td>
<td>EMK</td>
<td>voj¹</td>
<td>ka.pur¹, ka.pul¹ / xx</td>
<td>*kmboor</td>
</tr>
<tr>
<td>lose/lost</td>
<td>EMK, non-B</td>
<td>mot</td>
<td>bat³ / bat³</td>
<td>*pit</td>
</tr>
<tr>
<td>otter</td>
<td>EMK</td>
<td>raij³ ka³</td>
<td>pu.se³</td>
<td>*psaj-*phaj</td>
</tr>
<tr>
<td>python</td>
<td>EMK</td>
<td>ran³</td>
<td>pu.sign¹ loan⁴ / pu.sign¹ lian⁴</td>
<td>*talán</td>
</tr>
<tr>
<td>right side</td>
<td>EMK</td>
<td>fañ³, dam¹ (right hand)</td>
<td>toam⁴ / toam²</td>
<td><em>?atiñm-</em>?atam</td>
</tr>
<tr>
<td>squirrel</td>
<td>EMK</td>
<td>šaw²</td>
<td>ci.mok³ / xx</td>
<td>*proc⁵</td>
</tr>
<tr>
<td>tooth, canine</td>
<td>EMK</td>
<td>nañ¹ (canine)</td>
<td>ka.neñ¹ / ka.neñ¹</td>
<td>*kneñ (tooth)</td>
</tr>
<tr>
<td>blow</td>
<td>BVKa</td>
<td>t⁶oj³</td>
<td>t⁶urh¹,</td>
<td>xx</td>
</tr>
<tr>
<td>(blowpipe)</td>
<td></td>
<td>t⁶u̇j̇h¹ / t⁶u̇t¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gums</td>
<td>BVKa</td>
<td>laj⁴</td>
<td>līn² / ?līn² / xx</td>
<td>*lāŋ</td>
</tr>
<tr>
<td>hang up</td>
<td>BVKa</td>
<td>maok³</td>
<td>bok³ / xx</td>
<td>*mbak</td>
</tr>
<tr>
<td>hatch</td>
<td>BVKa</td>
<td>de⁵</td>
<td>ra.jo³</td>
<td>*ceh</td>
</tr>
<tr>
<td>intestines</td>
<td>BVKa</td>
<td>ruat⁴</td>
<td>roçe⁴ / ʒuac⁴</td>
<td>*roce</td>
</tr>
<tr>
<td>leech, forest</td>
<td>BVKa</td>
<td>vat³</td>
<td>plim¹</td>
<td>*pløem</td>
</tr>
<tr>
<td>mouth</td>
<td>BVKa</td>
<td>miñ⁴</td>
<td>kaŋ³ / kaŋ³</td>
<td>*kaañ</td>
</tr>
<tr>
<td>split (v.)</td>
<td>BVKa</td>
<td>be⁵</td>
<td>xx / peh¹</td>
<td>*paah</td>
</tr>
<tr>
<td>bio⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weevil</td>
<td>BVKa</td>
<td>mò⁴</td>
<td>muc³ / mœc⁴</td>
<td>*krmoɔt</td>
</tr>
<tr>
<td>armpit</td>
<td>VKa, EMK?</td>
<td>nac</td>
<td>li.βak³ / li.vak³</td>
<td>IImpaak (shoulder)</td>
</tr>
<tr>
<td>basket, winnowing</td>
<td>VKa</td>
<td>nouŋ¹</td>
<td>ta.don¹ / ta.don¹</td>
<td>*kdon</td>
</tr>
<tr>
<td>beak</td>
<td>VKa</td>
<td>mò⁵</td>
<td>tam.βoc³</td>
<td>*cròsh</td>
</tr>
<tr>
<td>beat</td>
<td>VKa</td>
<td>dañ³</td>
<td>pip³, təŋ¹ / taŋ³</td>
<td>*diŋ</td>
</tr>
<tr>
<td>Number</td>
<td>Term</td>
<td>Proto-Tai</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>-------------</td>
<td>------------------------------</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>branch</td>
<td>VKa ŋaŋ², ŋaŋ³</td>
<td>ta.ken³</td>
<td>*ŋkeŋ</td>
</tr>
<tr>
<td>74</td>
<td>brother</td>
<td>VKa ʔaŋ¹</td>
<td>m̥aŋ⁴ / m̥aŋ³, m̥aŋ¹</td>
<td>*ʔamiŋ</td>
</tr>
<tr>
<td>75</td>
<td>cage</td>
<td>VKa loun⁵, cuŋ²</td>
<td>klun¹</td>
<td>*taruŋ</td>
</tr>
<tr>
<td>76</td>
<td>crow (v.)</td>
<td>VKa yaŋ³</td>
<td>ta.kal¹, ti.kar¹</td>
<td>*tkar</td>
</tr>
<tr>
<td>77</td>
<td>duck</td>
<td>VKa vit⁴</td>
<td>βit⁴ / vit⁴</td>
<td>*ʔadaa</td>
</tr>
<tr>
<td>78</td>
<td>fall, drop</td>
<td>VKa do⁵</td>
<td>tugh¹ / toh¹</td>
<td>*sdoh</td>
</tr>
<tr>
<td>79</td>
<td>fireplace</td>
<td>VKa bep³</td>
<td>ta.peh¹ / ta.peh¹</td>
<td>*tpeh</td>
</tr>
<tr>
<td>80</td>
<td>flood</td>
<td>VKa lu⁴, lu⁶</td>
<td>blu⁴ / lu⁴</td>
<td>*luut</td>
</tr>
<tr>
<td>81</td>
<td>fly, blue-bottle</td>
<td>VKa nən⁴</td>
<td>?m.laŋ¹ / mə.laŋ¹</td>
<td>(P) li.laŋ</td>
</tr>
<tr>
<td>82</td>
<td>frog</td>
<td>VKa ec³</td>
<td>?a.kuə³ / kuak²</td>
<td>*ʔagut</td>
</tr>
<tr>
<td>83</td>
<td>heavy</td>
<td>VKa naŋ⁴</td>
<td>naŋ⁴, ŋaŋ⁴ / naŋ³</td>
<td>*nəŋ</td>
</tr>
<tr>
<td>84</td>
<td>insipid, tasteless</td>
<td>VKa naŋ⁴</td>
<td>?i.tah¹ / ?u.tah¹</td>
<td>*ʔatiaŋ</td>
</tr>
<tr>
<td>85</td>
<td>louse, body</td>
<td>VKa r̥n⁵</td>
<td>xx / briŋ³</td>
<td>*r̥n⁴r̥n</td>
</tr>
<tr>
<td>86</td>
<td>melt, drip</td>
<td>VKa r̥⁵, r̥³</td>
<td>ti.jəh¹ / xx</td>
<td>*joh</td>
</tr>
<tr>
<td>87</td>
<td>nest</td>
<td>VKa to³</td>
<td>xx / ŋo⁴, to⁴</td>
<td>*sooh</td>
</tr>
<tr>
<td>88</td>
<td>open (something)</td>
<td>VKa mə⁵</td>
<td>bə⁴</td>
<td>*ʔaah</td>
</tr>
<tr>
<td>89</td>
<td>peel, skin</td>
<td>VKa lut⁴</td>
<td>xx / lus⁴</td>
<td>*lak</td>
</tr>
<tr>
<td>90</td>
<td>remember</td>
<td>VKa na³</td>
<td>xx</td>
<td>(P) ʔa.ji, (B) sa.ji, sa.ʔi</td>
</tr>
<tr>
<td>91</td>
<td>remove skin</td>
<td>VKa lut⁴</td>
<td>xx / lus⁴</td>
<td>*luat</td>
</tr>
<tr>
<td>92</td>
<td>rice, sticky</td>
<td>VKa nep³</td>
<td>dəp³</td>
<td>*deep</td>
</tr>
<tr>
<td>93</td>
<td>rumor, ask</td>
<td>VKa, B? haŋ¹</td>
<td>haŋ¹ / xx</td>
<td>*haŋ</td>
</tr>
<tr>
<td></td>
<td>BVKa? (<em>ask</em> in redupl.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>seed, grain</td>
<td>VKa haŋ²</td>
<td>ŋok³ / ka.ŋaŋ³, kaʔ.ŋaŋ³</td>
<td>*kəoŋ</td>
</tr>
<tr>
<td>95</td>
<td>shine, bright</td>
<td>VKa saŋ³</td>
<td>ŋu.βaŋ³ / la.vaŋ³</td>
<td>*braŋ</td>
</tr>
<tr>
<td>96</td>
<td>shoulder</td>
<td>VKa vaj¹</td>
<td>ka.laŋ³ / ka.laŋ³</td>
<td>*crlaŋ</td>
</tr>
<tr>
<td></td>
<td>BVKa?</td>
<td></td>
<td></td>
<td>*ʔapaal</td>
</tr>
<tr>
<td>97</td>
<td>small, little</td>
<td>VKa, BVKa?</td>
<td>ŋi³</td>
<td>ŋi³</td>
</tr>
<tr>
<td>98</td>
<td>termite</td>
<td>VKa, B? moj³</td>
<td>ku.ʔaŋ⁴ / ku.mul¹⁴</td>
<td>*kmuar</td>
</tr>
<tr>
<td>99</td>
<td>thunder</td>
<td>VKa, BVKa?</td>
<td>šom³</td>
<td>krim⁴</td>
</tr>
<tr>
<td>100</td>
<td>tongue</td>
<td>VKa liə̆⁶</td>
<td>ʔăjh², ʔăjh² / ʔăjh¹</td>
<td><em>lsas~</em>lias (to lick)</td>
</tr>
</tbody>
</table>

References


The Vieto-Katuic hypothesis: lexical evidence


This paper aims to examine tone correspondences in Vietic languages, namely Vietnamese, Ruc and Arem, and attest Haudricourt’s tonogenesis hypothesis.

1. Haudricourt’s Tonogenesis hypothesis of Vietnamese
Haudricourt's Tonogenesis hypothesis of Vietnamese (1954) is well known. It not only put an end to the controversy on what language family Vietnamese belongs to, but also proposed a model of the development of tones in a language. Even after fifty years since its publication, it is generally accepted and still very influential among linguists. The hypothesis is usually explained as follows.

\[
\begin{array}{cccc}
 1 & 2 & 3 & 4 \\
 pa = pa = pa = ba & \text{paX} > \text{pá} > \text{p}á > \text{bá} & \text{pah} > \text{p}à > \text{p}á > \text{bá} \\
 ba = ba > pà > bá & \text{baX} > \text{bá} > \text{p}á > \text{b}á & \text{bah} > \text{b}à > \text{p}á > \text{b}á
\end{array}
\]

In stage 1, Vietnamese started as a toneless language. In stage 2, three pitch contours were formed depending on the coda types: level pitch from null coda, rising pitch from stop coda including glottal stop, and falling pitch from voiceless fricative coda. In stage 3, they split into higher and lower series according to [+/- voice] feature of onsets, and the number of tones doubled causing onsets of lower series devoiced. In stage 4, voiceless onsets were voiced again, though this stage is not relevant to the number of tones. This model may be called a consonantly-based model of tonogenesis.

However, simple questions arise. Tone is a phonological category, and it is realized and perceived by phonetic pitch. This pitch is encoded acoustically by fundamental frequency (F0), and this F0 cannot exist without voicing. If I accept the consonantly-based model, such syllables as [ai] or [au] without consonants cannot have tones. And specifically, how can a change from [bá] to [pá] between stage 2 and 3 explained? This change from voiced high rising pitch to voiceless low pitch does not seem to be plausible.

My question leads to a certain confidence that it is not consonants themselves, but interactions between various aspects of phonation that determines the pitch. This kind of idea is proposed by Thurgood (2001) in the name laryngeally-based model of tonogenesis.

1 This paper is the main result from my Master of Linguistics sub-thesis at the ANU (2004). I’m indebted to my supervisor Dr. Paul Sidwell for his enormous efforts, and lots of comments on it by Dr. Mark Alves and Dr. Phil Rose. I’m also indebted to Dr. Graham Thurgood for a new concept specifically applied in this paper.

2. Ruc and Arem – Vietic languages
I’m going to focus on the languages of Arem, Ruc and Vietnamese. The term “Vietic” is used to be known as “Viet-Muong”. Some scholars such as Ferlus prefer to use the latter name.

Population of Ruc and Arem is extremely small. According to the census in 1985, it is Arem 76, Ruc 125, Malièng 715, Mày 715, Sách 625 (Phong et al. 1988). Another source shows Chùt 2400 (VNA 1996). The name Chùt is used by the government to represent these extremely minor ethnic people settled in the North-Central area of Vietnam.

The survey of Ruc language started in 1986 by Russian and Vietnamese linguists (Loi 1993). However, the report is not available at hand. Available materials at hand are the following three.

(2) - Nguyễn Phú Phong, Trần Trí Dỗi and M.Ferlus (1988) on Ruc
- Nguyễn Văn Lợi (1993) on Ruc
- Michel Ferlus (1997, original version 1991) on Arem and Ruc

My study is based on these sources and not based on the field work of my own. In the SEALS Conference, Mark Alves and Michel Ferlus have contributed papers on Ruc language in 1997, 1999 and 2001. These papers are also cited.

3. Lexicostatistic Data
The following lexicostatistic data show lexical distance between the languages of Vietic and Katuic.

(3) Compared Vietic and Katuic lexica (Samarina, I.V. 1989, cited in Loi 1993)

<table>
<thead>
<tr>
<th></th>
<th>Bru</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T a o i h</td>
<td>63.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P a k o h</td>
<td>71.0</td>
<td>72.0</td>
</tr>
<tr>
<td>Ruc</td>
<td>Arem</td>
<td>40.0</td>
<td>38.0</td>
</tr>
<tr>
<td></td>
<td>66.5</td>
<td>37.5</td>
<td>33.0</td>
</tr>
<tr>
<td></td>
<td>53.0</td>
<td>51.5</td>
<td>35.5</td>
</tr>
<tr>
<td></td>
<td>50.5</td>
<td>50.0</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td>74.0</td>
<td>51.5</td>
<td>53.0</td>
</tr>
</tbody>
</table>

As is shown, Vietnamese shares nearly 50% common lexica with Ruc and Arem, while the rates are less than 30% with Katuic languages.

4. Tone Description
Tone description is crucially important but is a bit complicated because they are not shown by pitch alone but voice quality differences are involved in Vietic languages.
4-1. Standard Northern Vietnamese: 6 tones
Standard Northern Vietnamese is well known to have six tones, in Vietnamese name ngang, sác, họi, huyện, nàng, ngâ.

<table>
<thead>
<tr>
<th></th>
<th>(plain)</th>
<th>(abrupt)</th>
<th>(contour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ngang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>sác</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>họi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>huyện</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>nàng</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>ngâ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

They are grouped into historical higher and lower series, and each two tones make three pairs: plain tones, abrupt tones and contour tones. Besides differences in pitch, nàng tone is accompanied by glottal stop, and ngâ tone has creaky voice. Tone names in brackets, which I call comparative tone numbers, will be used to show tone correspondences more explicitly; the numbers (1 and 2) indicate higher and lower series, and the alphabets (A, B and C) indicate three contour types.

4.2 Ruc: 4 tones
Ruc is reported to have four tones.

1. High level pitch
2. Low level pitch (with breathy voice) or falling pitch (with breathy voice)
3. Rising pitch (with optional glottal stop)
4. Concave pitch or falling pitch (with optional glottal stop)

According to Loi (1993), tone 1 has high level pitch, and tone 2 has two variations in free variation, accompanied by breathy phonation. Tone 3, rising pitch can accompany glottal stop optionally. Tone 4 is reported to have two free variations, concave pitch and low falling pitch, and the falling pitch is accompanied by glottal stop optionally. Phong et al. (1988) describes six tones for Ruc, which counts tones 3 and 4 with and without glottal stop as an individual one.

4.3 Arem: Toneless
Arem is reported to be toneless, but to have breathy phonation, and post-vocalic laryngeal constriction (Ferlus 1997, 2001).

4.4 Research Questions
Here are some questions. According to Haudricourt, the tones developed from 0 > 3 > 6. How are they different from Ruc’s 4-tone system? What are the tone correspondences among Arem, Ruc and Viet?

5. Tone Developments – A tentative tree diagram
I propose a tentative tree diagram of Vietic from a viewpoint of the development of tones.
(6) *Proto-Vietic (toneless, contrastive phonation types)

Arem *Proto-RucViet (4 tones)

Ruc *Proto-VietMuong (6 tones)

South Viet, Central Viet (5 tones) Muong, N-C Viet (5 tones) North Viet (6 tones)

I assume the tones developed from 0 > 4 > 6, in the same way as Arem, Ruc and Viet. 5-tone systems of Southern, Central, North-Central varieties of Vietnamese as well as Muong are assumed to be the result of merger of two tones of the Proto-VietMuong, though categorisation of the two tones are not the same.

6. Word list

Word list is attached as an Appendix to this paper. I have found 109 cognate sets from the three materials and made them in a comparative list. On top from the left, after numbers and English glossary, there are transcriptions of Arem in Ferlus (1997), Ruc in Ferlus (1997), Ruc in Loi (1993) and Ruc in Phong et al. (1988) as in the original source. Next three columns are Vietnamese orthography and comparative tone numbers. In the last column, Ferlus’ 1997 reconstruction is shown as a reference. On left from the top, the 109 lexica are categorised into five groups by their coda types: (1) null coda, (2) fricative coda, (3) liquid coda, (4) stop codas, and (5) nasal codas. The colour further categorises the lexica within the coda types: coloured cells show basic regular correspondences, and colourless cells show irregular correspondences. From next section, I will demonstrate several typical examples of regular and irregular correspondences.

7. Regular Tone Correspondences (79/109 cases)

I will explain the regular tone correspondences according to the six tones of Standard Northern Vietnamese. This regular correspondence applies to 79 cases out of 109 lexica. The number after Ruc and Viet IPA transcription refers to its tone.

7.1 Ngang tone (1A)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V(Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>‘three’</td>
<td>pæ:</td>
<td>paː 1</td>
<td>baː 1A</td>
<td>paː</td>
</tr>
<tr>
<td>47</td>
<td>‘fly (verb)’</td>
<td>pal</td>
<td>pɔl 1</td>
<td>bǎːj 1A</td>
<td>pɔr</td>
</tr>
<tr>
<td>78</td>
<td>‘cooked rice’</td>
<td>kɔːm</td>
<td>kɔːm 1</td>
<td>kɔːm 1A</td>
<td>kɔːm</td>
</tr>
</tbody>
</table>

Vietnamese tone 1A corresponds to Ruc tone 1 in null coda, liquid coda and nasal coda syllables. Liquid coda in Arem and Ruc corresponds to the glide coda [j] in Vietnamese.
### 7.2 Sắc tone (1B) with null coda and sonorant codas

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V(Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>‘fish’</td>
<td>akä:?</td>
<td>aka: 3</td>
<td>ka: 1B</td>
<td>a-kä:?</td>
</tr>
<tr>
<td>55</td>
<td>‘roof’</td>
<td>(n.a.)</td>
<td>baël 3</td>
<td>mäj 1B</td>
<td>bäl?</td>
</tr>
<tr>
<td>92</td>
<td>‘four’</td>
<td>puän?</td>
<td>póm 3</td>
<td>bon 1B</td>
<td>póm?</td>
</tr>
</tbody>
</table>

Vietnamese tone 1B corresponds to Ruc tone 3 in the above mentioned three coda types, and final glottal stop in Arem. Ferlus (2001) indicates that this syllable-end glottal stop of Arem is used to represent “the constricted rimes in voiced ending”.

### 7.3 Huyễn tone (2A)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V(Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>‘fly (insect)’</td>
<td>uruæj</td>
<td>møræj 2</td>
<td>ruæj 2A</td>
<td>m-røj</td>
</tr>
<tr>
<td>53</td>
<td>‘long’</td>
<td>(n.a.)</td>
<td>jœal 2</td>
<td>zaœj 2A</td>
<td>ja:r</td>
</tr>
</tbody>
</table>

Vietnamese tone 2A corresponds to Ruc tone 2, and Arem transcription grave accent above vowels in most cases. According to Ferlus 2001, this diacritic indicates “breathy voice”.

### 7.4 Nằng tone (2B) with null coda and sonorant codas

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V(Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>‘wake up’</td>
<td>jïl?</td>
<td>jïl 4</td>
<td>zïj 2B</td>
<td>jïr?</td>
</tr>
<tr>
<td>108</td>
<td>‘heavy’</td>
<td>naŋ? (?)</td>
<td>naŋ 4</td>
<td>nãŋ 2B</td>
<td>naŋ?</td>
</tr>
</tbody>
</table>

Vietnamese tone 2B corresponds to Ruc tone 4, which corresponds to Arem combination of ‘breathy voice’ and ‘the constricted rimes in voiced ending’ in most cases.

### 7.5 Hội tone (1C) and ngã tone (2C)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V(Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>‘grass’</td>
<td>(n.a.)</td>
<td>kõh 1</td>
<td>kõ: 1C</td>
<td>kõh</td>
</tr>
<tr>
<td>44</td>
<td>‘nose’</td>
<td>müh</td>
<td>mu:rh 2</td>
<td>muj 2C</td>
<td>mus</td>
</tr>
</tbody>
</table>

Next sets show Vietnamese tones 1C and 2C, which correspond to Ruc tones 1 and 2 respectively whose rhymes have voiceless fricative [h], and which correspond to Arem voiceless fricative coda, too. In Arem and Ruc, final –h is still segmental and not tonal yet. In this point, the evidence is different from Haudricourt’s hypothesis.
7.6 Sắc tone (1B) and养老 tone (2B) with stop codas

Next sets show Vietnamese tones 1B and 2B with stop codas, which correspond to Ruc tone 1 and 2, or 3 and 4 respectively. What is interesting is the difference between Ferlus and Loi. While Ferlus describes 1 and 2, high level and low level, Loi describes 3 and 4, high rising and low falling. Since transcriptions of the two authors are coherent, I understand it is due to the perception difference.

7.7 Proposed Tonogenesis model
To summarise the above regular tone correspondences, proposed Tonogenesis model is shown as below.

Several remarks should be noted here. Comparative tone numbers are applied to Ruc, where Ruc tones 1, 2, 3 and 4 are shown as 1A, 2A, 1B and 2B respectively. The letter “a” represents all the vowels, the letter “n” represents nasal codas in /m, n, ŋ/, the letter “t” represents stop codas /p, t, k/. The apostrophe shows post-vocalic laryngeal constriction.
As we see above, Proto-Vietic is assumed to have contrasts in voice quality; modal vs. breathy, as Arem. Another laryngeal feature is the post-vocalic laryngeal constriction. This glottal constriction is what Diffloth (1989) called ‘creaky voice’. These two laryngeal configurations seem to be closely related to the origin of tones.

To express in binary features, breathiness and laryngeal constriction may be marked features. Following figures may be an evidence for this.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1A (ngang)</td>
<td>1136</td>
<td>1B (sắc 1)</td>
<td>891</td>
</tr>
<tr>
<td>2A (huyễn)</td>
<td>951</td>
<td>2B (nặng 1)</td>
<td>636</td>
</tr>
</tbody>
</table>

These figures show number of syllables by tones, which I counted from a modern Vietnamese-Chinese dictionary (Honda 2004), which lists 5795 syllables as a whole. According to this, 1B and 2B are less than 1A and 2A respectively due to the marked laryngeal constriction, and 2A and 2B are less than 1A and 1B respectively due to the marked breathiness.

The second important point is the presence of fricative coda /–h/ in Arem and Ruc, which is assigned to tone 1A in Ruc. I see chronological order of the emergence of tone 1C was much later than 1B.

The reason for my assumption is not only due to the presence of /–h/ in Ruc. In modern Vietnamese, we can easily find two more syllable types that are not shown in (13); that is nasal coda with tones 1C and 2C. There is no clarification why there are no Mon-Khmer cognates in this category (Gage 1985). They are assumed to be added by the influence of another language. In other words, emergence of contour tones might be triggered by the language contact, and that time the coda /–h/ worked as medium to give slots to contour tones. Arem and Ruc were exempted from this phonological change.

**8. Irregular Tone Correspondences (31/109 cases)**

Next, I will explain irregular correspondences. I posit three factors for the irregular correspondences, by which 17 cases out of 31 can be solved. The reasons for the rest 14 cases are unknown at present.

**8.1 Spreading of voicing from minor-syllable to main-syllable (7 cases)**

The first factor is spreading of voicing from minor-syllable to main-syllable. Before showing the examples, we should bear in mind the word structure of Arem and Ruc.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arem</td>
<td>CV+CCVC (?)</td>
<td>sesqui-syllable</td>
<td></td>
</tr>
<tr>
<td>Ruc</td>
<td>CV+CCVC / T</td>
<td>sesqui-syllable</td>
<td></td>
</tr>
<tr>
<td>Middle Vietnamese (17C)</td>
<td>CCVC / T</td>
<td>mono-syllable</td>
<td></td>
</tr>
<tr>
<td>Modern Vietnamese</td>
<td>CVC / T</td>
<td>mono-syllable</td>
<td></td>
</tr>
</tbody>
</table>

Word structure of Arem and Ruc are sesqui-syllable, which consists of minor-syllable and main-syllable. Examples of up-shifting and down-shifting are shown as below.
(16) Up-shift (#5, 6, 11, 104)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>‘deep’</td>
<td>ciru:</td>
<td>ʔu: 2A</td>
<td>sōw 1A</td>
</tr>
<tr>
<td>104</td>
<td>‘thunder’</td>
<td>kʰrimʔ</td>
<td>kʰrim 2B</td>
<td>sōm 1B</td>
</tr>
</tbody>
</table>

Above examples show tones are up-shifted due to the [-voice] feature of the minor-syllables. Although lower tones (2A, 2B) are expected, they are higher tones (1A, 1B) in Viet.

(17) Down-shift (#7, 36, 109)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>‘chicken’</td>
<td>lakæ:</td>
<td>rəkaːʔtəka: 1A</td>
<td>ʔa: 2A</td>
</tr>
<tr>
<td>36</td>
<td>‘husked rice’</td>
<td>ðəkəʔtəkə: 1B</td>
<td>yaːw 2B</td>
<td></td>
</tr>
</tbody>
</table>

Above examples show tones are down-shifted due to the [+voice] feature of the minor-syllables. Although higher tones (1A, 1B) are expected, they are lower tones (2A, 2B) in Viet.

8.2 Misperception of phonation type difference with voicing of onsets and tones (4 cases)
The second factor is misperception of phonation type difference with voicing of onsets, which affected tones.

(18) Up-shift (#12, 13, 87, 88)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V (Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>‘go’</td>
<td>ti:</td>
<td>ti: 2A</td>
<td>di: 1A</td>
<td>di: ? ti:</td>
</tr>
<tr>
<td>87</td>
<td>‘bone’</td>
<td>səŋ</td>
<td>saːŋ 2A</td>
<td>swəŋ 1A</td>
<td>dʒaŋ ? tʃaŋ</td>
</tr>
</tbody>
</table>

In both cases, although lower tone (2A) is expected, they are higher tone (1A). Ferlus’ reconstruction also suggests two options: voiced and voiceless onsets. I posit breathy voice because I have a similar experience in transcribing East Javanese. Examples below show sub-minimal pair.

(19) East Javanese

<table>
<thead>
<tr>
<th>Gloss.</th>
<th>*Mistranscription</th>
<th>Corrected transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘four’</td>
<td>*pa’pa’t’ +l</td>
<td>pa’pat’ +l</td>
</tr>
<tr>
<td>‘father’</td>
<td>*pʰa’paʔ +l</td>
<td>pa’paʔ +l</td>
</tr>
</tbody>
</table>

At first, I transcribed the first syllables as contrastive [-aspiration] and [+aspiration] on the initial consonant. The informant insisted they are ‘voiceless’ and ‘voiced’. However, both initial consonants are apparently voiceless. After repetitions of reproduction, I found the difference is rather in phonation type on the following vowels: the contrast between [modal voice] and [breathy voice], or modal voice and slack voice (Ladefoged and Maddieson 1996: 63). The latter is also accompanied by a slightly low pitch. I suppose similar sounds existed and similar misperception took place in the history of Vietnamese.
8.3 Semantic change and innovations (6 cases)
The third factor is semantic change and innovations. They are shown in pairs. One example is as below.

(20) Semantic change and innovations (#73/74, 99/100, 105/106)

<table>
<thead>
<tr>
<th>No.</th>
<th>Gloss.</th>
<th>Arem</th>
<th>Ruc</th>
<th>Viet</th>
<th>*P-V (Ferlus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>‘eye’</td>
<td>mät</td>
<td>mảt 1B/2B</td>
<td>mất 1B</td>
<td>mat</td>
</tr>
<tr>
<td>74</td>
<td>‘face’</td>
<td>(n.a.)</td>
<td>mảt 1B/2B</td>
<td>mất 2B</td>
<td>mat</td>
</tr>
</tbody>
</table>

Although Vietnamese ‘eye’ and ‘face’ is phonologically different, it seems they have been derived from one etymon.

9. Summary of comparison – Haudricourt’s model vs. Vietic evidence
Above analysis is summarised as follows:

- Phonation type difference (modal vs. breathy) is consistently reflected within Vietic.
- Chronological order of 2-way tone split by initial consonants to higher and lower series is not placed in the third stage of the development of tones, but the distinction between modal vs. breathy voice and the accompanying pitch distinction, had supposedly existed since the early stage of the proto language.
- Compared glottal stop [-ʔ] with fricative [-h] in syllable-end position, glottal stop was incorporated to tone system much earlier, while glottal fricative was incorporated to tone system later, as Arem and Ruc evidence shows.
- Therefore, the tone development was not 0 > 3 > 6 as Haudricourt, but 0 > 4 > 6, or even 0 > 2 > 4 > 6 may be possible, though evidence of 2-tone system is not found yet.
- Three main factors for irregular correspondences are posited: (1) spreading of voicing from minor-syllable to main-syllable, (2) misperception of phonation type difference with voicing of onsets and tones, and (3) semantic change and innovations.

10. Toward a laryngeally-based model of Tonogenesis
Lastly, I present a chart of the laryngeally-based model of tonogenesis below.
As it shows, the acoustic energy to create pitch difference was not directly supplied from consonants. It is the phonation, or voice quality, originally derived from initial and final consonants, which is the overlapping portion between consonants and vowel, that worked as medium or reservoir of acoustic energy. In the case of Vietic, this voice quality is both breathy voice and post-vocalic laryngeal constriction. Each of them becomes responsible for pitch height and pitch contour respectively, when the pitch element becomes more salient than voice quality, that is when the system changes from register system to tone system.

References
Alves, Mark (1997) “Ruc and other minor Vietic languages: linguistic strands between Vietnamese and the rest of the Mon-Khmer language family”, papers from SEALS 7th conference.
### Tone correspondences and tonogenesis in Vietic

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Arem</th>
<th>Ruc</th>
<th>Rue</th>
<th>Rue</th>
<th>Rue</th>
<th>Mod VN</th>
<th>Exp. Tone</th>
<th>*P-VM</th>
<th>Ferlus 1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>three</td>
<td>pæː</td>
<td>pa¹</td>
<td>pa¹</td>
<td>pa</td>
<td>ba</td>
<td>1A</td>
<td>1A</td>
<td>pa:</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>drunk</td>
<td>pɔɾːt</td>
<td>pʰɾit²</td>
<td>pri¹</td>
<td>pri pʰʊdo</td>
<td>sɑy</td>
<td>1A</td>
<td>1A</td>
<td>pri: &gt; pʰɾit / kʰɾit:</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>arm, hand</td>
<td>tʰː</td>
<td>sːi¹</td>
<td>sːi¹</td>
<td>sːi</td>
<td>tay</td>
<td>1A</td>
<td>1A</td>
<td>sːi</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ear</td>
<td>tʰːaj</td>
<td>sːaj¹</td>
<td>sːaj¹</td>
<td>sːaj</td>
<td>tain</td>
<td>1A</td>
<td>1A</td>
<td>sːaj</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>deep</td>
<td>ɕɪɾːɕ</td>
<td>tːu³</td>
<td>chɔɾu¹, djɔɾu¹, tru¹, tu¹</td>
<td>rʰpʲ / ʈuːɲ kʰsàŋ</td>
<td>sʔau</td>
<td>1A</td>
<td>2A</td>
<td>cruc: &gt; kʰruː</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>buffalo, carabao</td>
<td>(n.a.)</td>
<td>cɑlu²</td>
<td>tlu¹, klu¹</td>
<td>klu</td>
<td>tɾâu</td>
<td>1A</td>
<td>2A</td>
<td>elu: &gt; klu: / tlu:</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>chicken (gen.)</td>
<td>lakaː</td>
<td>tːa ka¹</td>
<td>roka¹, toka¹</td>
<td>rʰkɑ</td>
<td>gɑ</td>
<td>2A</td>
<td>1A</td>
<td>rka:</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>betel</td>
<td>uliːw</td>
<td>(plu²) S</td>
<td>plu²</td>
<td>plu</td>
<td>trạːu, giạːu</td>
<td>2A</td>
<td>2A</td>
<td>blu: &gt; blu: / tlu:</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>sky, heaven, God</td>
<td>tlæj¹ (?)</td>
<td>plaj²</td>
<td>ploj²</td>
<td>plaj</td>
<td>trọːi, giọːi</td>
<td>2A</td>
<td>2A</td>
<td>b-laj &gt; blaj / tlaj</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>fly (insect)</td>
<td>ur uproj¹</td>
<td>morooj², porooj²</td>
<td>mǔroj</td>
<td>ruői</td>
<td>2A</td>
<td>2A</td>
<td>m-ruːj</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rain</td>
<td>mːa</td>
<td>kumɔː²</td>
<td>kumɔː²</td>
<td>kʰmɔː</td>
<td>mua</td>
<td>1A</td>
<td>2A</td>
<td>k-ma: &gt; kʰm:ma</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>go</td>
<td>fː</td>
<td>ti²</td>
<td>ti²</td>
<td>ti</td>
<td>ḏːi</td>
<td>1A</td>
<td>2A</td>
<td>d i / ti</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>tail</td>
<td>tːuaj</td>
<td>(tːuaj²)</td>
<td>tːuoj²</td>
<td>tuəj</td>
<td>duːi</td>
<td>1A</td>
<td>2A</td>
<td>dzɛj / təj</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>lip</td>
<td>(n.a.)</td>
<td>cəboj¹</td>
<td>churbøj³</td>
<td>củbôːj</td>
<td>mői</td>
<td>1A</td>
<td>1B</td>
<td>c-βuːj / c-βuj</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>bundle</td>
<td>pɔːʔ</td>
<td>pɔːc</td>
<td>(n.a.)</td>
<td>bɔ</td>
<td>1B</td>
<td>1B</td>
<td>pɔʔ?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>blood</td>
<td>mːawʔ</td>
<td>təmːu³</td>
<td>[ʔaːsam']</td>
<td>[ʔaːsɾːm]</td>
<td>mái</td>
<td>1B</td>
<td>1B</td>
<td>t-muː? &gt; t'muː?</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>clothes</td>
<td>ʔəwʔ</td>
<td>ʔəw³</td>
<td>ʔəw³</td>
<td>ʔəw</td>
<td>ʔo</td>
<td>1B</td>
<td>1B</td>
<td>ʔawʔ?</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>dog</td>
<td>aːcːʔ</td>
<td>aːcːʔ</td>
<td>aːko³</td>
<td>aːko³</td>
<td>aːko³</td>
<td>1B</td>
<td>1B</td>
<td>aːko³?</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>female</td>
<td>kɛːʔ</td>
<td>pokːi³</td>
<td>pːkiʔ</td>
<td>gài</td>
<td>1B</td>
<td>1B</td>
<td>kɛʔ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>fish</td>
<td>aːkːɛʔ</td>
<td>aːkːa³</td>
<td>aːkːa³</td>
<td>aːkːa³</td>
<td>cǎ</td>
<td>1B</td>
<td>1B</td>
<td>ʔa-kǎʔ</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>have</td>
<td>kɔʔ</td>
<td>kɔ³</td>
<td>kɔ³</td>
<td>kɔ³</td>
<td>cø</td>
<td>1B</td>
<td>1B</td>
<td>kɔʔ?</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>leaf</td>
<td>uləːʔ</td>
<td>uləː¹</td>
<td>ula³, hla³</td>
<td>ula³</td>
<td>lá</td>
<td>1B</td>
<td>1B</td>
<td>s-ləʔ?</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>louse, head</td>
<td>ctːʔ</td>
<td>cːi³</td>
<td>cːi³</td>
<td>cːi</td>
<td>cʰɑː, chʰi</td>
<td>1B</td>
<td>1B</td>
<td>cːʔ?</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>paper</td>
<td>kɛcːjʔ</td>
<td>kɛcːj³</td>
<td>kʰcːj³</td>
<td>kʰcːj</td>
<td>giːy</td>
<td>1B</td>
<td>1B</td>
<td>k-ɛcːj?</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>remember, think about</td>
<td>ɲaʔ</td>
<td>kapa³</td>
<td>kunhːo³</td>
<td>kʰpa</td>
<td>nʰɔ</td>
<td>1B</td>
<td>1B</td>
<td>k-ɲaʔ? &gt; kʰɲaʔ?</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>salt</td>
<td>boj³</td>
<td>boj³</td>
<td>boj³</td>
<td>boj</td>
<td>mʊi</td>
<td>1B</td>
<td>1B</td>
<td>fɒjʔ</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>six</td>
<td>pʰɾawʔ</td>
<td>faw⁴</td>
<td>pɾaw³, psaw³, phɾaw³, saw³</td>
<td>fɾaw</td>
<td>sʰau</td>
<td>1B</td>
<td>1B</td>
<td>p-ruːʔ &gt; pʰruːʔ</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>--------------------</td>
<td>------</td>
<td>------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>smoke</td>
<td>ahù́?</td>
<td>kahx1</td>
<td>kahho1</td>
<td>kahx2</td>
<td>khoi</td>
<td>1B</td>
<td>1B</td>
<td>k-hx2j &gt; k=bxj? / t=bxj?</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>wind (blowing)</td>
<td>kajù?</td>
<td>kajox2</td>
<td>kajox3</td>
<td>giòi</td>
<td>1B</td>
<td>1B</td>
<td>k-jox2?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>bee, wild honeybee</td>
<td>kwí</td>
<td>kwí3</td>
<td>(n.a.)</td>
<td>khoài</td>
<td>1B</td>
<td>1B</td>
<td>kwe?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>rice (plant and grain)</td>
<td>alac?</td>
<td>alo3</td>
<td>álô</td>
<td>lúa</td>
<td>1B</td>
<td>1B</td>
<td>?-lx?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>fruit</td>
<td>ulec?</td>
<td>puli3</td>
<td>păli</td>
<td>trăi</td>
<td>1B</td>
<td>1B</td>
<td>p-le? &gt; p-le? / tle?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>stone</td>
<td>atax?</td>
<td>lata1</td>
<td>lata1, ta1</td>
<td>ta?</td>
<td>1B</td>
<td>1B</td>
<td>l-ta?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>forest</td>
<td>braw?</td>
<td>bru1</td>
<td>bru3</td>
<td>brú</td>
<td>[rùng] rú</td>
<td>1B</td>
<td>m-rú?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>mother</td>
<td>(n.a.)</td>
<td>mie1</td>
<td>mèe4</td>
<td>me, me1</td>
<td>2B</td>
<td>2B</td>
<td>me?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>rice (husked)</td>
<td>nôkx?</td>
<td>tako1</td>
<td>rokó1</td>
<td>gao</td>
<td>2B</td>
<td>1B</td>
<td>r-ko?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>break</td>
<td>peh</td>
<td>peh1</td>
<td>(n.a.)</td>
<td>peh</td>
<td>bé, bé</td>
<td>1C</td>
<td>1C</td>
<td>peh, peh</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>grass</td>
<td>(n.a.)</td>
<td>koh1</td>
<td>koh1</td>
<td>có</td>
<td>1C</td>
<td>1C</td>
<td>koh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>red</td>
<td>(t=s)N, tsh1</td>
<td>tuko2</td>
<td>to?</td>
<td>dó</td>
<td>1C</td>
<td>1C</td>
<td>tsh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>cloth</td>
<td>kupal? (?)</td>
<td>(kupal?) N</td>
<td>kupal1</td>
<td>vai</td>
<td>1C</td>
<td>1C</td>
<td>kpaas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>vomit</td>
<td>abah</td>
<td>bah1</td>
<td>(n.a.)</td>
<td>máa</td>
<td>1C</td>
<td>1C</td>
<td>båh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>breathe</td>
<td>anòh</td>
<td>tangoh1</td>
<td>tàmph</td>
<td>thò, ngàri</td>
<td>1C</td>
<td>1C</td>
<td>tàns &gt; t’ñaw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>nest</td>
<td>nùh</td>
<td>suh1</td>
<td>lò?</td>
<td>[o] [to]</td>
<td>1C</td>
<td>1C</td>
<td>suh / ?uh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>seven</td>
<td>pah</td>
<td>(paj?) N</td>
<td>paj4</td>
<td>bây, bây</td>
<td>1C</td>
<td>1C</td>
<td>pas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>nose</td>
<td>müh</td>
<td>murh1, murh2</td>
<td>muh, muh1</td>
<td>mîi</td>
<td>2C</td>
<td>2C</td>
<td>mus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>tongue</td>
<td>fish</td>
<td>laar1</td>
<td>loarh1, loarh2</td>
<td>laarh, laarl</td>
<td>laòri</td>
<td>2C</td>
<td>lacs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>fly (verb)</td>
<td>pal</td>
<td>pal1</td>
<td>por1</td>
<td>pal</td>
<td>bay</td>
<td>1A</td>
<td>1A</td>
<td>par</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>lime (substance)</td>
<td>apul</td>
<td>kapul1</td>
<td>kapul1</td>
<td>kāpul</td>
<td>vöi</td>
<td>1A</td>
<td>1A</td>
<td>kpur</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>two</td>
<td>hîrél</td>
<td>hal1</td>
<td>hal1</td>
<td>hâi</td>
<td>1A</td>
<td>1A</td>
<td>hâr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>cloud</td>
<td>(n.a.)</td>
<td>mol1</td>
<td>mol1</td>
<td>mòl</td>
<td>1A</td>
<td>1A</td>
<td>-mol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>tree</td>
<td>(n.a.)</td>
<td>(kaj?)</td>
<td>koj2</td>
<td>cày</td>
<td>1A</td>
<td>1A</td>
<td>kôl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>return</td>
<td>(n.a.)</td>
<td>(vi’)</td>
<td>[Br?]</td>
<td>(n.a.)</td>
<td>vé</td>
<td>2A</td>
<td>2A</td>
<td>ve’l</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>long</td>
<td>(n.a.)</td>
<td>jâl1</td>
<td>jàl1</td>
<td>?jâl</td>
<td>dâî</td>
<td>2A</td>
<td>2A</td>
<td>jär</td>
<td></td>
</tr>
</tbody>
</table>
### Tone correspondences and tonogenesis in Vietic

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>plow, plough</td>
<td>(n.a.)</td>
<td>kãal²</td>
<td>koTAIL²</td>
<td>kãl</td>
<td>cãy</td>
<td>2A</td>
<td>2A</td>
<td>gal</td>
</tr>
<tr>
<td>55</td>
<td>roof</td>
<td>(n.a.)</td>
<td>ba³</td>
<td>bãl`</td>
<td>bãl</td>
<td>mãl</td>
<td>1B</td>
<td>1B</td>
<td>bãl²</td>
</tr>
<tr>
<td>56</td>
<td>egg</td>
<td>[uluʔ?]</td>
<td>[tul³]</td>
<td>[tul³, tul³]</td>
<td>trãg`</td>
<td>1B</td>
<td>1B</td>
<td>thu?</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>wake, get up, rise</td>
<td>ji³</td>
<td>ji³</td>
<td>(n.a.)</td>
<td>pãp jìri</td>
<td>dãy</td>
<td>2B</td>
<td>2B</td>
<td>jìri?</td>
</tr>
<tr>
<td>58</td>
<td>rise, raise</td>
<td>aj³</td>
<td>aj³</td>
<td>(n.a.)</td>
<td>(n.a.)</td>
<td>dãy</td>
<td>1B</td>
<td>1B</td>
<td>?a-jàr?</td>
</tr>
<tr>
<td>59</td>
<td>run</td>
<td>(n.a.)</td>
<td>tãngaj³</td>
<td>tãpøj</td>
<td>chay</td>
<td>2B</td>
<td>1B</td>
<td>jàl?</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>light (weight)</td>
<td>(n.a.)</td>
<td>neg³</td>
<td>nhe³</td>
<td>ŋnél</td>
<td>nhe</td>
<td>2B</td>
<td>1B</td>
<td>neg³</td>
</tr>
<tr>
<td>61</td>
<td>dust</td>
<td>(n.a.)</td>
<td>(n.a.)</td>
<td>[pœ⁴]</td>
<td>[kãij, kãjãj]</td>
<td>bui`</td>
<td>2B</td>
<td>2B</td>
<td>buiž?</td>
</tr>
</tbody>
</table>

**[Final -p, -t, -c, -k]**

| 62  | lightning | cœ³p | cœ³p¹ | [alar²] | cœp | chãp` | 1B | 1B | cœp |
| 63  | sing | ahæt | hat³ | hat³ | hât | 1B | 1B | hat |
| 64  | iron | (n.a.) | k³kat³ | khat³ | klãt | sãt` | 1B | 1B | krac > k³rac |
| 65  | cool | (n.a.) | ramace³ | mát | mát | 1B | 1B | tmace |
| 66  | cut | kac³ | kac³ | [pœm²] | kãc | cãt, chãt | 1B | 1B | kãc |
| 67  | sand | tkac | tkac | tkac⁴ | tkãc | cãt | 1B | 1B | tkac |
| 68  | hair (head) | ut³uk | usuk³ | usuk³ | usûk` | tôc` | 1B | 1B | suk |
| 69  | stump, base of tree | tk³ok | tk³ok¹ | [kul koaj] | tôkôk koaj | gõc` | 1B | 1B | tkôk |
| 70  | water | dæk` | dák³ | dák³ | nœc` | 1B | 1B | dák |
| 71  | one | mœc` | mœc⁴ | móc` | mœc` | 2B | 2B | móc |
| 72  | wear | (mêk) | mãk³ | mãk³ | mãk`ለw | mã`c | 2B | 2B | mãk |
| 73  | eye | mít` | mít³ | mít³ | mát` | 1B | 2B | mat |
| 74  | face | (n.a.) | (n.a.) | (n.a.) | (n.a.) | mát` | 2B | 2B | mat |

**[Final -m, -n, -p, -ŋ, -m?, -n?, -ŋ?, -ŋ?]**

<p>| 75  | bird | icim<code> | icim¹ | ʔcim</code> | ʔcim | chim<code> | 1A | 1A | cim</code> |
| 76  | five | dam<code> | dam¹ | ñdam¹ | ñdãm</code> | nãm<code> | 1A | 1A | ñdãm</code> |
| 77  | hundred | tlam<code> | klãm¹ | klãm</code> | trãm<code> | 1A | 1A | klãm / klãm | | 78  | rice (cooked) | ko⁴m | ko⁴m¹ | (chaw</code>) | ko⁴m | com<code> | 1A | 1A | ko⁴m | | 79  | year | (n.a.) | nam¹ | nam¹ | nãm</code> | nãm<code> | 1A | 1A | enom</code> |</p>
<table>
<thead>
<tr>
<th>80</th>
<th>child</th>
<th>ku⁴m</th>
<th>ku⁴m¹</th>
<th>ku⁴m`, kon⁴</th>
<th>ku⁴m`</th>
<th>com`</th>
<th>1A</th>
<th>1A</th>
<th>ko⁴m`</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>eat</td>
<td>?aŋ</td>
<td>?aŋ¹</td>
<td>?aŋ</td>
<td>ān</td>
<td>1A</td>
<td>1A</td>
<td>ān</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>flower</td>
<td>(n.a.)</td>
<td>pọŋ¹</td>
<td>pông¹, pIAL</td>
<td>bông</td>
<td>1A</td>
<td>1A</td>
<td>pọŋ</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>in</td>
<td>tkàn</td>
<td>klàŋ¹</td>
<td>klàŋ</td>
<td>inng</td>
<td>1A</td>
<td>1A</td>
<td>klàŋ &gt; klàŋ / tkàn</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>on, on troop</td>
<td>lnàn</td>
<td>alìŋ¹, lnàŋ¹</td>
<td>lnàn, lnàŋ¹</td>
<td>inên, lèn</td>
<td>1A</td>
<td>1A</td>
<td>klàŋ &gt; klàŋ / tkàn</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>tooth</td>
<td>ataŋ¹</td>
<td>kasàŋ¹</td>
<td>kásaŋ</td>
<td>râng</td>
<td>1A</td>
<td>1A</td>
<td>kásaŋ</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>ask, question</td>
<td>(n.a.)</td>
<td>hàn¹</td>
<td>(n.a.)</td>
<td>fhojài hàn</td>
<td>1A</td>
<td>1A</td>
<td>hàn</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>bone</td>
<td>sàŋ¹</td>
<td>sâŋ¹</td>
<td>sâŋ</td>
<td>xuông</td>
<td>1A</td>
<td>2A</td>
<td>dʒån / tʃiŋ</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>foot, leg</td>
<td>ciŋ¹</td>
<td>chîŋ²</td>
<td>ciŋ</td>
<td>chûn</td>
<td>1A</td>
<td>2A</td>
<td>ciŋ / ciŋ</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>bed</td>
<td>(n.a.)</td>
<td>kàciŋ²</td>
<td>ciŋ</td>
<td>guîŋîng</td>
<td>2A</td>
<td>2A</td>
<td>kjçengineering</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>near</td>
<td>(n.a.)</td>
<td>tkîŋ¹</td>
<td>cûkîŋ²</td>
<td>cûkgîng</td>
<td>1B</td>
<td>1B</td>
<td>tkîŋ</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>eight</td>
<td>tʰæːm?</td>
<td>(tʰæːm?) N</td>
<td>thâm², tohâm¹</td>
<td>thâm</td>
<td>1B</td>
<td>1B</td>
<td>sɑam?</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>four</td>
<td>pɔɔn?</td>
<td>pɔɔn¹</td>
<td>pɔn</td>
<td>bûn</td>
<td>1B</td>
<td>1B</td>
<td>pɔɔn</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>nine</td>
<td>cîn²</td>
<td>cîn¹</td>
<td>cîn</td>
<td>chîn</td>
<td>1B</td>
<td>1B</td>
<td>cîn?</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>ripe, cooked</td>
<td>cîn²</td>
<td>cîn¹</td>
<td>cîn</td>
<td>chîn</td>
<td>1B</td>
<td>1B</td>
<td>cîn?</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>shoot</td>
<td>pijî¹</td>
<td>pij¹</td>
<td>pij</td>
<td>bûn</td>
<td>1B</td>
<td>1B</td>
<td>pij?</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>snake</td>
<td>uʈʰîŋ?</td>
<td>pɔsîŋ³</td>
<td>pɔsîŋ</td>
<td>râŋ²</td>
<td>1B</td>
<td>1B</td>
<td>pɔsîŋ?</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>palm (of hand)</td>
<td>adîŋ¹</td>
<td>(n.a.)</td>
<td>(n.a.)</td>
<td>nâŋ</td>
<td>1B</td>
<td>1B</td>
<td>kûâŋ?</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>month</td>
<td>tʰæːŋ?</td>
<td>thâŋ²</td>
<td>thâŋ</td>
<td>kàrâŋ?</td>
<td>1B</td>
<td>1B</td>
<td>kàrâŋ?</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>wing</td>
<td>(n.a.)</td>
<td>kên¹</td>
<td>kên</td>
<td>cînh</td>
<td>1B</td>
<td>1B</td>
<td>kên?</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>branch</td>
<td>kùŋ? (V¹)</td>
<td>tkàŋ¹</td>
<td>tãkûŋ</td>
<td>cûnh</td>
<td>2A</td>
<td>1B</td>
<td>têŋ? / cêŋ?</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>bitter</td>
<td>(n.a.)</td>
<td>tàn¹</td>
<td>tûn³</td>
<td>ëtûn</td>
<td>1B</td>
<td>1B</td>
<td>tân</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>live, be alive</td>
<td>tloŋ¹</td>
<td>klûŋ¹, klûŋ⁴</td>
<td>kûn, kûnûn</td>
<td>sôn²</td>
<td>1B</td>
<td>1B</td>
<td>k-kûnûn / k-uôn?</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>salty</td>
<td>mîn²</td>
<td>mân¹</td>
<td>mân</td>
<td>mân</td>
<td>2B</td>
<td>2B</td>
<td>mân?</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>thunder</td>
<td>kûrîm²</td>
<td>kûrm¹</td>
<td>kûrîm / kûrm⁴</td>
<td>sâm²</td>
<td>1B</td>
<td>2B</td>
<td>kûrm?</td>
<td>k'ûrm?</td>
</tr>
<tr>
<td>105</td>
<td>stand</td>
<td>tiŋ? (³)</td>
<td>tûŋ²</td>
<td>tiŋ</td>
<td>dûng</td>
<td>1B</td>
<td>2B</td>
<td>tiŋ</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>build</td>
<td>potîŋ²</td>
<td>(n.a.)</td>
<td>pûtîŋ</td>
<td>dûng</td>
<td>2B</td>
<td>2B</td>
<td>p-tîŋ?</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>borrow</td>
<td>maŋ² (³)</td>
<td>moaŋ¹</td>
<td>moaŋ</td>
<td>moûn</td>
<td>2B</td>
<td>2B</td>
<td>maŋ?</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>heavy</td>
<td>naŋ? (³)</td>
<td>naŋ¹, ?naŋ⁴</td>
<td>nàŋ</td>
<td>nâŋ²</td>
<td>2B</td>
<td>2B</td>
<td>nàŋ?</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>louse (body)</td>
<td>&quot;rîŋ?&quot;, &quot;tîŋ?&quot;</td>
<td>brîŋ¹</td>
<td>brîŋ</td>
<td>rûn</td>
<td>2B</td>
<td>1B</td>
<td>mûn? &gt; p-rîŋ?</td>
<td></td>
</tr>
</tbody>
</table>
Tone correspondences and tonogenesis in Vietic

Remarks:

(?) : borrowing, noted by Ferlus  
[   ] : suspicious cognates, noted by Honda  
( ) S : Sach (alternative), noted by Ferlus  
( ) N : Nguôn (alternative), noted by Ferlus

Sources for Appendix:

Ferlus, Michel (1991, 1997 modified version) "Vocalisme du Proto Viet-Muong", Paper to be circulated at the 24th International Conference on Sino-Tibetan Languages and linguistics


Summary
The Katuic languages are a branch of the Mon-Khmer (MK) family spoken by more than one million people living in Thailand, Cambodia, Laos and Vietnam. While the majority of Katuic speakers live in eastern Thailand and Cambodia, the greatest diversity of Katuic languages lies in the Salavan and Sekong provinces of Laos and adjacent border areas of Vietnam, part of a complex patchwork of small ethnic communities. From a comparative-historical point of view Katuic has particular importance, as between them the languages appear to have conserved some very ancient phonological and lexical features. At the same time some Katuic languages have been remarkably innovative and developed some of the richest vowel systems in the world. The recent advances in the reconstruction of Proto Katuic (Sidwell 2005) potentially allow us to investigate the sub-grouping of Katuic within Mon-Khmer on the basis of comparative phonology. However, the results are somewhat ambiguous, and do not support any special sub-grouping of Katuic within Mon-Khmer.

Classification of Katuic with the Mon-Khmer family
During the first major phase of comparative-historical work on the MK languages, which lasted into the 1960s (effectively beginning with the work of Schmidt (1901, 1904, 1905 etc.) until Pinnow (1959) and Shafer (1952, 1965)), there was no coherent account of the real extent and internal structure of the Mon-Khmer family.

Thomas and Headley (1970) established a new paradigm when they successfully applied lexicostatistics to the emerging body of new field data, distinguishing nine branches: Pearic, Khmer, Bahnaric, Katuic, Khmuic, Monic, Palaungic, Khasi and Viet-Muong. Adding Aslian and Nicobarese (not examined by Thomas and Headley although already long recognised as MK), Diffloth’s (1974) expanded listing became the received classification. The Munda languages of India are also generally recognised as related to MK, although opinion is divided over how close that relationship is. All together they are recognised as forming the Austroasiatic phylum, but in this paper I am only concerned with analysis up to the MK level.

1 I would like to thanks Mark Alves for valuable advice and comment on this paper, in addition to his ongoing cooperation and encouragement in respect of my broader Mon-Khmer research activities. This paper was also made possible by assistance from the Department of Linguistics of the Max Planck Institute (Leipzig) in the form of stipend support during 2005.

2 Since the 1980s some minor languages of China have come to light that may or may not constitute a new branch.
There is no yet general agreement among specialists concerning how the MK branches further sub-group with each other. Peiros (1998) also applying lexicostistics, identified five coordinate MK groups (plus Munda):

1. Central (Bahnaric, Katuic, Aslian, Monic)
2. Vietic
3. Northern (Palaung-Wa, Khmuic)
4. Khmer
5. Khasi
6. Munda

In contrast Diffloth has developed a model, based largely on identifying lexical innovations (lexicostistics counts only retentions, at least as practiced by Peiros), that distinguishes only three coordinate MK groups (plus Munda), which in their most recent incarnation (2005) are designated “Khasi-Khamuic”, “Khmero-Vietic” and “Nico-Monic” (see Figure 2).

One notes that the placement of Katuic is radically different in each the above schema: Peiros grouping it with Bahnaric, Aslian & Monic, while Diffloth sub-groups it with Vietic (Viet-Muong). Alves (this volume) also cites lexical evidence in favour of Diffloth’s hypothesis. However, it is striking that two different lexical approaches, one based upon identifying retentions (Peiros) and the other innovations (Diffloth) reach such contrary results. In these circumstances a fresh approach is called for. I have recently completed a comprehensive phonological reconstruction of Proto-Katuic
Proto-Katuic phonology and the sub-grouping of MK languages

(Sidwell 2005), and a preliminary phonological reconstruction of Proto-Mon-Khmer is also available (Shorto f.c.), so an attempt at a classification based upon historical phonology is possible. Diffloth (1991) did invoke some phonological data in respect of the claim of Vieto-Katuic subgrouping, so this will be discussed first before moving on to presenting my historical-phonological analysis.

![Sub-grouping of Mon-Khmer (and Austroasiatic) languages according to Diffloth (2005)](image)

**Diffloth’s Vieto-Katuic hypothesis**

Diffloth (1991) proposes that in some cases, conditioned by unknown factors, prevocalic */h/ became */h* in Proto-Vieto-Katuic, and subsequently */s* in Vietic. The changes are attested in a small number of well distributed MK etyma, e.g.:
It is argued that phonetically it is more likely for Vietic *s to have come from a Proto-Katuic-Vietic *h, rather than independently from PEMK *ʔ. However, the potential counter-examples to this rule outnumber the examples, e.g. there was no such change in etyma such as PK *kaʔaʔ ‘crow’, *ʔŋʔuur ‘wasp’, *sʔaap ‘yawn’, *ʔuus ‘fire, firewood’ and others which have solid MK etymologies. In the absence of a motivated phonetic change it is very difficult to explain this h:s correspondence. My own hypothesis is that it resulted from some cluster reduction, but in what language and at what stage? The very limited lexical attestations of the change may rather indicate that the affected words were borrowed, perhaps a case of dialect mixing or a language shift. Borrowings from one or more related language can create small odd clusters of correspondences that stand out as in this case. On balance I suggest that this h:s correspondence is simply too ambiguous in its significance to base a classification upon it, and instead I suggest that the strongest weight should be placed upon clearly motivated systemic changes. This requires a well developed model of the historical phonology.

The Proto-Katuic consonants and PMK

The comparative phonological reconstruction of Proto Katuic (Sidwell 2005) reveals a mostly conservative system. The proto-consonant inventory is equivalent to what Shorto (f.c.) posits for PMK, with the addition that a palatal implosive is indicated, which I provisionally reconstruct to PMK—it appears that Shorto did not have extensive Katuic sources and missed the distinction.

<table>
<thead>
<tr>
<th>Khmer</th>
<th>Katuic</th>
<th>Vietic</th>
</tr>
</thead>
<tbody>
<tr>
<td>chʔɤɤ ‘bone’</td>
<td>Katu ʔhaap</td>
<td>Ruc saap, Viet. xuong</td>
</tr>
<tr>
<td>kʔaep ‘centipede’</td>
<td>Katu kahiŋ</td>
<td>Maleng kasiip</td>
</tr>
</tbody>
</table>

It is significant that Katuic retains the PMK three stop series intact—in fact it survives unaltered in various Katu dialects, but is reduced to two series by the merger of plain voiced and voiceless stops in the rest of Katuic. Outside of Katuic the old three series distinction is maintained in Bahnar, and in the orthography of Old Mon (see Shorto 1971 Introduction for discussion). In most MK languages it has reduced to two
series: the merger of plain voiced and voiceless stops that occurred in Most of Katuic is also typical of Northern MK and Vietic, while merger of the plain voiced and implosive stops into a single voiceless series occurred in Aslian and most of Bahnaric. However, the fate of the three stop series in Vietic is somewhat problematic. While it is clear that there was a devoicing of proto plain voiced stops in Vietic, leading to a register split (and ultimately the Vietnamese tones) it is not straightforward to reconstruct the voicing contrast to the Proto Vietic level. Strict application of the comparative method forces one to reconstruct a small number of exceptions with voiced initials, leading Feulner (1998) to suggest that Vietic formed by the mixing of two MK dialects, one of which had already undergone a devoicing (perhaps a Northern MK language)—but I shall ignore such exceptions for now and concentrate on the most solidly regular phonological correspondences.

In various languages PMK palatal stops (*c and even secondarily devoiced *ɟ) have merged with the reflexes of *s. Also there have been various mergers of *s and *h. Depending upon the sequence the above changes, sub-groups differ in their correspondence patterns among these segments, even though they may have a full set of palatals and fricative corresponding superficially to the PMK inventory.

Therefore, we can potentially gain some insight into the way PMK broke up by looking at the patterning of correspondences among the initial stops and fricatives in CVC words. The apparently regular correspondences and reconstructed PMK values are summarised as follows3:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>*p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>*p</td>
<td>p/p</td>
<td>ph</td>
<td>p/p</td>
<td>ph~p</td>
</tr>
<tr>
<td>*ɓ</td>
<td>b</td>
<td>b</td>
<td>p</td>
<td>p</td>
<td>b</td>
<td>b/m</td>
<td>b</td>
<td>b/p</td>
<td>p</td>
</tr>
<tr>
<td>*t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t/d</td>
<td>t</td>
<td>t/t</td>
<td>th~t</td>
</tr>
<tr>
<td>*ɗ</td>
<td>d</td>
<td>d</td>
<td>r</td>
<td>t</td>
<td>d</td>
<td>d/n</td>
<td>d</td>
<td>d/d</td>
<td>d</td>
</tr>
<tr>
<td>*d</td>
<td>d</td>
<td>d</td>
<td>t</td>
<td>t</td>
<td>d</td>
<td>d/th</td>
<td>t</td>
<td>t/t</td>
<td>t</td>
</tr>
<tr>
<td>*c</td>
<td>c</td>
<td>c</td>
<td>s</td>
<td>ch</td>
<td>c</td>
<td>c/e</td>
<td>s</td>
<td>s/s</td>
<td>s</td>
</tr>
<tr>
<td>*ʄ</td>
<td>ʄ</td>
<td>ʄ</td>
<td>(?)</td>
<td>c</td>
<td>ʄ</td>
<td>ʄ/c</td>
<td>ʄ</td>
<td>ʄ/c</td>
<td>(?)</td>
</tr>
<tr>
<td>*j</td>
<td>j</td>
<td>j</td>
<td>(?-c)</td>
<td>j</td>
<td>j</td>
<td>j/c</td>
<td>j</td>
<td>j/c</td>
<td>(?)</td>
</tr>
<tr>
<td>*s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s/th</td>
<td>h</td>
<td>h/h</td>
<td>f</td>
</tr>
<tr>
<td>*k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k/k</td>
<td>k</td>
<td>k/k</td>
<td>kh~k</td>
</tr>
<tr>
<td>*g</td>
<td>g</td>
<td>g</td>
<td>k</td>
<td>g</td>
<td>g</td>
<td>g/k</td>
<td>g</td>
<td>g/k</td>
<td>g</td>
</tr>
<tr>
<td>*h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h/h</td>
<td>h</td>
<td>h/h</td>
<td>h</td>
</tr>
<tr>
<td>*ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ/ʔ</td>
<td>ʔ</td>
<td>ʔ/ʔ</td>
<td>ʔ</td>
</tr>
</tbody>
</table>

Table 3: Correspondences of MK initial stops

First of all some remarks about my reconstruction of the PMK *ʄ/*j distinction—this is based upon my reconstruction of the same in Proto-Katuic, a

3 The table is based upon the equivalent in Shorto (f.c.), modified to account for subsequent advances in reconstruction, e.g. Shorto, writing in the 1970s, had no access to adequate Vietic or Katuic reconstructions.

197
tendency in Bahnar to devoice reflexes of *ɟ but not *ʄ, and the failure of *ʄ to merge with /s/ in Pearic (although the latter is based upon on a couple of examples, and the Pear consonants otherwise pattern differently to Katuic and Bahnaric).

If we group the branches according to their developments in the stops and fricatives the closest matches with Katuic are Bahnaric and Monic, particularly in respect of the retention of imploded stops. We also note that while Bahnar has /s/ from *c, this is a Central Bahnaric innovation and Proto-Bahnaric *c is indicated. Vietic, suggested by Diffloth (2005) and Alves (this volume) as sub-grouping with Katuic, is not excluded—it is evident that there was a general devoicing of stops in (or before) Proto-Vietic, and a retention of the imploded stops as plain voiced stops, and reversing these changes reveals a system directly corresponding to Old Mon.

However, the retention of the imploded series may not be a good indicator of sub-grouping. It is apparent that the loss of implosion and consequent merger with other series has occurred independently at time in MK. For example, within Bahnaric it is only (apparently) Bahnar that retains the etymological imploded series, while others have merged them with the voiced stops. The northern languages (Khasi, Palaungic, Khmuic) pattern neatly in respect of the palatals and fricatives (confirming this sub-grouping as suggested by both Diffloth 1974 and Ferlus 1974), and while most of the languages lost contrastive implosion, distinct reflexes of the imploded series remain in, for example, Khasi and the Riang stops, indicating that the PMK consonantism survived intact into Proto-Northern-Mon-Khmer (after the northern lenition of palatals began).

It is also clear that various devoicings of stops occurred independently. Generally devoicing of stops in MK is an areal trend shared with the intrusive Tai languages, while those located out of contact (such as Aslian), or attested before Tai contact (Old Khmer, Old Mon) do not show it. We are therefore left with no apparent evidence among the stop correspondences for distinguishing any sub-grouping among the non-northern languages.

The Proto-Katuic vowels and PMK

The comparative phonological reconstruction of Proto Katuic (Sidwell 2005) reveals a conservative system that most likely lacked tones or voice quality distinctions. I reconstruct the Proto Katuic vowel inventory as follows:

<table>
<thead>
<tr>
<th>*i</th>
<th>*ɨ</th>
<th>*u</th>
<th>*ii</th>
<th>*ii</th>
<th>*uu</th>
<th>*ie</th>
<th>*iə</th>
<th>*uo</th>
</tr>
</thead>
<tbody>
<tr>
<td>*e</td>
<td>*ə</td>
<td>*o</td>
<td>*ee</td>
<td>*oə</td>
<td>*oo</td>
<td>*ia</td>
<td>*ia</td>
<td>*ua</td>
</tr>
<tr>
<td>*ε</td>
<td>*a</td>
<td>*ɛ</td>
<td>*ɛɛ</td>
<td>*aa</td>
<td>*ɔɔ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: *Proto Katuic vowel inventory

The reconstruction of PMK vowels is more problematic. Shorto (1975, ms.) suggested the following based upon the set of regular correspondences between Mon and Khmer:
Proto-Katuic phonology and the sub-grouping of MK languages

<table>
<thead>
<tr>
<th>*i</th>
<th>*u</th>
<th>*ii</th>
<th>*uu</th>
<th>*iə</th>
<th>*ai</th>
<th>*uə</th>
</tr>
</thead>
<tbody>
<tr>
<td>*e</td>
<td>*ə</td>
<td>*ə</td>
<td>*ee</td>
<td>*əə</td>
<td>*oo</td>
<td></td>
</tr>
<tr>
<td>*a</td>
<td>*ə</td>
<td>*ə</td>
<td>*aa</td>
<td>*əə</td>
<td>*cc</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Shorto’s Proto Mon-Khmer vowel inventory

At this time I am not able to present a comprehensive tabulation of MK vowels as I have done for the consonants above—even Shorto (f.c.) does not attempt it, instead basing his vocalic reconstruction upon Old Mon and Khmer (see Shorto 1976 for discussion). In the discussion that follows I will simply try to demonstrate that underlyingly Proto-Katuic vocalism is readily derived from PMK. The innovations I have reconstructed are not evident in any other MK branch, although further research may well reveal more relevant facts.

Shorto’s PMK *ai alternates occasionally with *aa, so I suspect that it reflects a special development of *aa—it is not discussed further here. PK *ia and *ua often reflect Shorto’s PMK *iə and *uə although this is not always the case (the picture is complicated by the fact that some of Shorto’s *iə and *uə reconstructions are undetermined and may actually reflect PMK *ii and *uu or some perhaps even diphthonged PMK *ɛɛ and *ɔɔ). The lack of *ɛɛ or *ɛ in Shorto’s reconstruction is not definitive—front vowels are less frequent than central or back vowels in MK (and some other SEA families) and the shortage of examples makes it difficult to demonstrate a mid versus low front vowel contrast at the PMK level.

Given the above considerations, it appears that PK or its immediate ancestor innovated the diphthongs *ie, *uə, *iə, *ia and the monophthong *ii. Comparison between PK forms with these vowels and their wider MK cognates may shed light on on this question.

The PK diphthongs *ie, *uə were specifically reconstructed to account for correspondences that were identical to *ii and *uu except that the reflexes in Pacoh are tense vowels [o0o] and [ɛɛ] instead of the regular reflexes of *ii and *uu. It may also be that my PK *ie, *uə are artifacts, and that there was a conditioned split in Pacoh, but I doubt this as there is no indication that the Pacoh registers arose with a general split in the vowel system (vis. Khmer) but rather a vowel shift (vis. North Bahnaric).

Below are comparisons between relevant PK forms with these diphthongs and MK etymologies extracted from Shorto f.c. Note that for brevity I have not included semantic values for the various MK daughter reflexes.

Examples of PK *uə and *ie with MK etymologies/reconstructions:

**PK *muoj ‘one’**
Sh. PMK 1495. muuj, muaj ‘one’
Cf. Khmer muuj, Chrau muaj, Khmu mooj

**PK 1292 *puol ‘prophesise’**
Sh. PMK 1753. *pul, *puul, *pual ‘to divine, take an augary’
Cf. Khmer pool, Chrau pool, Viet. bọi, [Chamic: Roglai pual]
PK 1294. *ʔasuom ‘shrimp’
Sh. PMK 1419a. *sum, *suum ‘shrimp, prawn’
Cf. Viet. thực

PK 1299. *huor, *hool ‘sing with flame’
Sh. PMK 1685. *hur ‘to expose to heat’
Cf. Stieng, Srê, Bahnar *nuur
Sh. PMK 1812. *[r]huul, *[r]ha[ ]l ‘to grill’
Cf. Biat Stieng nhool, Golar Bahnar babaal, Khmer rool

PK 765. *ʔieh , *ʃieh ‘sew’
PMK 1897. *[j]is, *[j]ias, *[j]as ‘prick’
Cf. Car Nicobar iiciih-[h]ata, L Mon jeh, Bahnar jah, Khmer jas, Semang ćos

PK 767. *cieh ‘cut open’
PMK 1982. *cih, *ciash ‘to split’
Cf. Lawa saių, Palaung seh, Viet. chê, Bahnar ceh

PK 779. *ʔariew ‘bell’
PMK 1870. *krhiow ‘small bell, jingle’
Cf. Mon hareų, Khasi [sha] kuhiaw

The results of these comparisons are mixed, and perhaps say as much about the limitations of Shorto’s vowel reconstruction. In many cases MK reflexes do not allow one to distinguish clearly between a monophthong or a diphthong in PMK, hence Shorto offers variant reconstructions that he called “alternances” e.g. *uu ~ *ua, *ii ~ *iə. Underlying these is a hypothesis (see Shorto 1976) that the “alternances” existed in the proto-languages (perhaps expressing derivational or other functions). Unfortunately there is no clear patterning between Katuic and Shorto’s PMK vocalism that allows us to distinguish a deeper origin for PK *uo, *ie—and I suspect that they are basically Katuic internal innovations resulting from diphthongisation of high vowels.

More interesting are PK diphthongs *iə, *ia and *iː—where it appears that wider comparisons are able to shed some light on the matter.

Generally speaking, it appears on my reading of the data that Shorto was correct in not reconstructing PMK high central vowels—where they do occur in MK languages it as apparent that they are secondary, arising from diphthonging and/or raising of low and/or mid-vowels, and less frequently from centering of high vowels.

The raising of low central vowels is very common in MK, and may occur several times in the history of a given language (or sub-group)—and this is particularly evident in the case of Katuic.

I found that for Katuic there are many comparisons that unambiguously indicate *iː and *i—where the modern reflexes show few conditioned changes so I would not expect any dispute about these. External comparisons generally suggest that they reflect raised allophones of PMK *əə and *ə. By contrast the reconstruction of PK *iə and *ia are less straightforward, and a summary of the supporting correspondences follows:
Proto-Katuic phonology and the sub-grouping of MK languages

<table>
<thead>
<tr>
<th>PK</th>
<th>Kui</th>
<th>Bru</th>
<th>Ta’Oi</th>
<th>Kriang</th>
<th>Katu</th>
<th>Pacoh</th>
</tr>
</thead>
<tbody>
<tr>
<td>*ɨə</td>
<td>-</td>
<td>ɨə-ɔa</td>
<td>ia</td>
<td>ia</td>
<td>ia</td>
<td>oo /_m</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>ia-ɔa</td>
<td>ee</td>
<td>ii</td>
<td>ia</td>
<td>oo /_r</td>
</tr>
<tr>
<td>*ɨa</td>
<td>ii/ia</td>
<td>ia/ia</td>
<td>ii</td>
<td>ii-ee</td>
<td>ii-ia</td>
<td>ia</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>ia/ia</td>
<td>ii</td>
<td>ii</td>
<td>ia</td>
<td>ia /_/</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>ia</td>
<td>ia</td>
<td>ia</td>
<td>/_c</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>ia/ia</td>
<td>ii</td>
<td>ii</td>
<td>ia</td>
<td>ia</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>ia/ia</td>
<td>ia</td>
<td>ia</td>
<td>/_/</td>
</tr>
</tbody>
</table>

Table 6: Correspondences supporting PK *ɨə and *ɨa

As can be seen in the above table, the distinction between *ɨə and *ɨa is based upon contrasting reflexes in Pacoh, and (like *uo, *ie) could be disputed, but that is not the important point here because I am arguing that in any case *ɨə and *ɨa, whether they are one or two phonemes in PK, are ultimately secondary, and derived from PMK *aa. In some case the raising is very old, in other cases it is more recent. The latter is seen in the reconstruction of PK *aa/*ɨa and *aa/*ɨə doublets. All examples of *aa/*ɨə go back to unambiguous PMK *aa, while the one case of *aa/*ɨə occurs in a Katuic lexical innovation (‘husband’). E.g.:

**PK 88. *kajaal, *kajial ‘wind’**
PMK 1782. *kjaal ‘wind’

**PK 38. *ksaac, *siac ‘scoop/splash (water)’**
PMK 872. *sac, *saac ‘to bale out’

**PK 85. *ʔhaal, *ʔghial ‘light weight’**
PMK 1801. *saal ‘light weight’

**PK 78. *kajaak, *kajiel ‘husband’**

Additionally there are etyma with solid MK etymologies that support the reconstruction of PK *ɨə and *ɨa, suggesting a much earlier raising. E.g.:

**PK 481. *ʔatəm, *ʔatəm ‘right side’**
PMK 1353. *st1am, *st1uum ‘right-hand’

**PK 482. *dian ‘stretch’**
PMK 544a. *taan ‘to extend, stretch’
PK 484. *dias ‘break’
PK 501. *miat ‘vulture’
PK 487. *crlia ‘thorn’
PK 759. *priit, *priat ‘banana’

Mostly the wider MK reflexes of these (and other relevant) etyma have low vowel reflexes or their conditioned variants, in other words there are no other MK branches that share these changes in this lexicon. The one exception I have found is the parallel in Bahnaric *priit ‘banana’, although an isolated example like this could easily reflect borrowing, and the fact that Katuic reflexes support the reconstruction of *priit, *priat doublet forms supports this.

Conclusion
In so far as I can offer any sort of conclusion at this time, my research in progress on Katuic and wider Mon-Khmer historical phonology confirms the NorthernMK subgrouping of Khasi, Palaungic and Khmuic that has longstanding acceptance by scholars. On the other hand, I have found no particular phonological evidence to support subgrouping Katuic with any other branch, or even to sub-group any of the non-northern languages.

This suggests a basically flat tree structure along the following lines (leaving aside the issue of Munda), ordered roughly from north to south:

- Northern (Khasi, Palaungic, Khmuic)
- Vietic
- Katuic
- Bahnaric
- Khmer
- Pearsic
- Monic
- Aslian
- Nicobarese

Figure 3: Mon-Khmer classification suggested by comparative phonology

From this point one turn to lexical methods to discover any evidence of sub-grouping among the non-northern branches, and in this matter I refer readers to the paper by Alves in this volume. Alves’ conclusions are not inconsistent with Diffloth’s
Proto-Katuic phonology and the sub-grouping of MK languages

model, as outlined in Figure 2, that effectively treats the non-northern branches as forming a sub-group in opposition to NMK.

References