CHAPTER 4

METHODOLOGY

4.0 Introduction

Chapter 4 describes the purpose and goals of this research project, as well as the wordlist and sociolinguistic questionnaire collection procedures and data sources.

4.1 Purpose and goals

The purpose of this survey is to determine a likely reference dialect[6] among the selected Tai Nua varieties. To achieve this purpose, the following goals were selected.

1) To investigate how closely related the selected speech varieties appear to be to one another;

2) To investigate language attitudes to determine if the selected Tai Nua dialects are likely to be maintained to the next generation; and

3) To discover if there is a variety of Tai Nua that is respected and reportedly understood sufficiently for communication among the other Tai Nua speech varieties.

In this thesis, there is some information from other areas provided for a broader context, especially in Myanmar. The analysis, therefore, also presents the lexical similarity results between Tai Nua in these areas.

4.2 Wordlist Collection and Procedures

A 550 item wordlist, developed by Bryan Allen, of SIL East Asia Group, was used in this survey. Wordlists were collected and transcribed using the International Phonetic Alphabet (IPA). The wordlists were recorded using a mini-disc recorder. These sessions normally took place in Tai villages in houses, restaurants and temples. Transcribing and making audio recordings of the words on the list took a few hours for each participant.

For lexical comparison, 100 items were selected from 550 item wordlist. These 100 items are those suggested by Mann (2004). Mann's comparative wordlist complies the 281 and 436 item wordlists from SIL used in Mainland Southeast Asia with Matisoff's 209 item CALMSEA wordlist and the Swadesh 100 and 200 item wordlists. The words are ranked by how often they appear on the various wordlists. Items from the wordlists collected followed the order prescribed by Mann (2004) unless there are apparent semantic differences in certain sites or no lexical forms are given in more than 3-4 sites. In these cases the lexical set is skipped and the next set is used in its place.

These 100 items were compared between each word pair for different speech forms to determine how many items were the same or very similar. Decisions about similarity grouping were made using a set of criteria for determining lexical similarity as mentioned in Figure 18 in Chapter 3. Then the number of matches in each category is determined if the word pair meets the minimum lexical similarity criteria mentioned in Figure 19 in chapter 3. If this criteria is met, the word pair is considered lexically similar.

Microsoft Excel was employed to keep track of the lexical similarity decisions for all word pairs. Those word pairs which were considered lexically similar were given a value of '1', while those word pairs considered not lexically similar were

---

[6] LinguaLinks 1996-1999 defines a reference dialect as "one variety of a language chosen to be developed into the standard, written form."

given a value of '0'. A few basic functions in Excel were used to calculate the lexical similarity percentages between each pair of speech varieties. In order to summarize the results, the percentages of lexical similarity are organized in a table. Finally, the "Unweighed Pairs Grouped Method with Arithmetic Average" method, or UPGMA (J. Grimes 1995:69) and programs in the suite of PHYLIP 3.6 (Feldsenstein 2002), were used to obtain a lexicostatistic tree diagram.

Site selections were based on the known distribution of Tai Nua speakers. In China, the research was done in Yunnan Province. The selected sites are Yingjiang (No. 1), Baoshan (No. 2), Fengqing (No. 3), Ruili (No. 4), Zhefang (No. 5), Mangshi (No. 6), Lincang (No. 7), Mengting (No. 8), Gengma (No. 9), Shuangjiang (No. 10), Canyuan (No. 11), Jinggu (No. 12), Siamo (No. 13) and Jinghong (No. 14) as shown in Figure 39.

Figure 39: Sites selected of Tai Nua speech varieties in Yunnan, China.

Data for the Tai Ya language was provided by Person (2005). There are data for Tai Mao and Tai Lai in Myanmar from 6 different sites. The data for Tai Mao is from the Northern Shan State and the data for Tai Lai is from Khatcho, Ketda, Ywait, Homalin, and Maungkham.

4.2.1 Data sources

Sanders (1980:26) suggested that a poor informant choice may provide non-representative lexical data leading to erroneous calculations. Therefore, screening informants is an important process which should not be ignored. To do this the researcher should collect background information from the informant to screen the wordlist candidate before collecting a wordlist. This background information includes name, gender, age, occupation, education level, mother tongue, place of birth and duration of current residence. If the person is not a native speaker of the local variety of interest, another informant should be sought.

In this thesis the wordlists were collected from male Tai Nua native speakers. Most of these informants were between 40 and 60 years of age. The survey sought to keep age relatively constant to assure the compatibility of the wordlist data.

4.3 Sociolinguistic Questionnaires

There are two kinds of questions employed in questionnaires: open and closed. Open questions will give freedom to interviewees to present their comments, but by doing this, they could stray from the focus of the research. An example of this type of question is: "Describe your reactions to this speaker" after the interviewee has heard an audio sample of a speaker. In a closed question, the interviewee will be asked yes-no, multiple choice, or ranking systems questions. This type of question is easier for the researcher to score although it will force the interviewee to answer the question in the researcher's terms. Therefore the ideal survey design would be to use open questions to conduct the initial research and closed questions based on the answers to the open questions to support the final results

(Fasold 1984:145-147). Larson (2002) suggests that follow-up questions such as 'why?', 'who?', and 'when?' are helpful after the main questions.

This questionnaire used in this research employs both open and closed questions as well as follow-up questions.

4.3.1 Questionnaire content

There are 30 questions in the questionnaire used in this research. The questionnaire is shown in Chinese in appendix 1 and translated into English in Appendix 2. The questionnaire begins with background information: age, gender, occupation, education level, birth place, and reported locations where they have lived. This demographic information gathering is not included in the 30 question total.

Generally, there are five areas of research focus in this questionnaire. The first part explores language identity. Language use in the area and demographic population are considered under this part because they are all related with the name of the language. This section consists of 15 questions about the identity of the people: questions #1-15.

The second part explores language use. This part asks when, where, and with whom people use a language or languages (Larson 2002). It shows which languages people are comfortable using in certain situations. Fifteen questions (all listed under question number 16) in the questionnaire investigate 15 domains to identify which languages people choose to use.

The third part explores dialect variation and perceived intelligibility. The goal of these questions is to discover which speech varieties are perceived to be the same, which a little different, and which very different from the language of the person being interviewed. There are 5 questions in this part, excluding a few follow-up questions associated with each main question: questions # 17-21.

The fourth part explores language attitudes of Tai Nua speakers toward their own language and other languages. This part also covers language vitality because the continued use of a language is strongly dependent on language attitudes. There are eight questions that cover language use among the children and the interest of the local people toward a potential written form: questions #21-30.

The fifth part explores Tai Nua self-reported Chinese bilingualism. Each question is associated with a proficiency level based on the FSI[7] (Foreign Service Institute of the United States Department of State) scale. There are 15 sub-questions under question number 31 in this section.

4.3.2 Questionnaire administration procedure

Five Chinese-English speakers were employed as translators. They were trained on how to elicit information using the questionnaire with the researcher and sometimes by themselves. The objectives of the survey were explained to them so that they would understand how to record the answers. The researcher or translator wrote down the interviewee's responses. It took about 15-20 minutes to complete the questionnaire. When the researcher completed fieldwork, all the data was input into Microsoft EXCEL.

To get responses from several villages in each survey site, the research team often visited markets as people from many places gather there. Schools, temples, village centers and fields were often visited as well. Fasold (1984) described a sample as follows:

[7] Now referred to as ILR (Interagency Language Roundtable)

> A sample consists of a small number of members of a population which can be studied in detail. The results can then be projected to the population as a whole. In order for this projection to be accurate, the sample should be a microcosm of the whole population...Making sure that the sample represents the population in all crucial ways is not easy (1984:86-87).

The sample used in this research is a quota sample, i.e. where a number of respondents from different categories are interviewed. These categories include people from different ages, genders and education levels. This type of sample allows for the consideration of trends within categories and, assuming that a composite of these categories reflect the wider community, trends in the general population. The interviewees were selected to satisfy the categories shown in Table 9.

Gender	Men			Women		
Age	15-30	31-45	46 up	15-30	31-45	46 up
No education						
Primary						
Secondary						
College						

Table 9: A chart of the sample (adapted from Blair 1990:38)

In Table 9, "No education" refers to those who have never been in school or went to school for less than 3 years; "Primary" refers 3 to 6 years in school; "Secondary" 6 to 12 years in school; "College" refers to those who finish the university after graduating from high school.

Blair (1990:36-39) suggests that the minimum number of interviews for each cell shown in Table 9 (such as Women, age 15-30 with no education) should be 5. However, in some situations the number may need to be adjusted based on the demographics, for instance, there may not be women over 46 with college education. In these situations it is acceptable that the researcher found few or no interviewees in these cells. For example, in this thesis no women over 30 years old with a college degree were found. The number and distribution of the interviewees is discussed in section 4.3.4.

4.3.3 Limitation of the questionnaires

The questionnaire has limitations in three areas. These areas are language identity or group name, language perception and bilingualism.

In the language identity section, many of the interviewees did not know how to answer the questions. For example, they were often confused about language names, both what other people call their language and what they themselves call their language.

In the language perceptions section, people from the same site often gave different answers. For example, in Shuangjiang, two male interviewees were individually asked "where the language is different from their village but they still can mostly understand." One of the interviewees replied that Tai in Mengting was intelligible, while another interviewee said Tai in Mengting replied that it was so different from his language that he could not understand most of them. This difference in responses makes it difficult for the researcher to make a firm statement about comprehension. However, the decisions in this research were based on trends from the majority of interviewees.

In the bilingualism section, there are five levels of self reported proficiency which correspond to the FSI scale from 1 (minimum proficiency) to 5 (native speaker proficiency). Those interviewees who are not able to pass lower level proficiency should not be able to pass upper level proficiency. However some of them are not

sure how fluent they are in the second language, so they were reluctant to give an answer to the researcher. This makes it hard for the researcher to categorize them precisely. Since this is reported information, it is possible that there is variation in the interviewee's perceptions of their bilingual ability. In some cases, the interviewees did not complete this section for a variety of reasons, especially since it was the last one on the questionnaire. There were 16 subjects who did not finish answering every question. These subjects were excluded. So the data analyzed is from the 135 subjects who completed the questionnaire.

4.3.4 Data sources

The goal was to test different types of informants using Table 9. However many of the Tai in China, especially women, do not have the opportunity to study in secondary or upper level education. Thus, it was almost impossible to find people in these categories. The research team attempted to interview both men and women. Of the sociolinguistic interviewees 81 (54%) were men and 70 interviewees (46%) were women.

Male	Female
81	70

Table 10: The gender of the sample

Table 11 shows the categories of people interviewed in this research. The total number of people in this sample is 151.

Gender	Men			Women			Total	Percent of total
Age	15-30	31-45	46 up	15-30	31-45	46 up		
No education	3	15	15	9	7	15	64	42%
Primary	12	12	7	15	9	4	59	39%
Secondary	6	2	3	8	2	0	21	14%
College	2	2	2	1	0	0	7	5%
Total	23	31	27	33	18	19	151	100%

Table 11: The Sample

In the sample shown in Table 11, uneducated interviewees provided 42 percent of the data and the educated interviewees (primary school to college) provided 58 percent. Thus, there is a slight bias toward interviewees with some education in this sample.

Table 11 categorizes the occupations of the interviewees.

Category Number	Occupations	Total
1	Government officers (including teachers)	9
2	Students	16
3	Farmers	70
4	Employees (such as construction workers, factory workers, car watchers)	7
5	Monks	8
6	Restaurant workers (owners, waiters, waitresses, food sellers)	18
7	Others (including housewife)	23

Table 12: Occupations

Since most of the Dai are farmers, 70 people of the interviewees described themselves as farmers. Some of them do not work or care much about making money especially in the small villages. However if they live in or move to a bigger city, where there are other minorities found, like Ruili or Simao, usually they will either work in a restaurant as a waiter or sell food in the market.

Table 13 shows the number of the interviewees in each site.

Location	Number
Mangshi	10
Ruili	10
Yingjiang	10
Zhefang	10
Shuangjiang	10
Gengma	13
Mengting	8
Canyuan	19
Lincang	13
Fengqing	17
Baoshan	11
Jinhong	0
Simao	10
Jinggu	10
Total	151

Table 13: The selected locations

In Jinghong, the research team decided not to administer any questionnaires when they were informed that there was only one Tai Nua village called 'Mak Huai Tong' in the area. The rest of the people in this area are Tai Lue speakers. Therefore only an informal open interview with one villager was conducted. The analysis of Tai Nua attitudes in Jinghong are reported in chapter 6.