Chapter 2
Theoretical foundations

This chapter describes the theoretical foundations for the methodology used in this study. The relatedness of the selected varieties is assessed by phonetic and lexical comparison. Mutual intelligibility is measured by Recorded Text Testing. These are described in Sections 2.1, 2.2 and 2.3. The sociolinguistics investigation emphasizes bilingualism, language choice, language vitality and language attitudes. These topics are briefly discussed in Sec. 2.4.

2.1 Phonetic and lexical comparison

This section describes the wordlist used to collect lexical items and the procedure of comparing them to determine the percentage of lexical similarity.

2.1.1 Wordlists

Wordlists are commonly used for basic language survey. A wordlist can help surveyor's efficiency in analyzing a language. A certain amount of data can also be collected in a limited time in a survey situation. Relevant word lists should be chosen to the language area being surveyed.

Swadesh (1952, 1955) suggested a list of 100 words representing 'core vocabulary' that should be relevant for all languages. Mann (2004) compared various wordlists that have been used in Southeast Asia including universal lists such as Swadesh (1955) and lists that claim to be culturally relevant to language family of either the wider Asian region or mainland Southeast Asia in particular. The total number on all the lists amounted to 504 items (Mann 2004:25). By combining similar wordlists to avoid biasing the result, Mann counted how many lists contained each item to arrive at a ranking of the 504 items. The items ranked highest were those items that are contained in several of the different wordlists. Mann proposed that the higher ranked items be given priority when comparing languages of the region.

The wordlist used in this thesis began as the SIL MSEA 281-item wordlist, a list based on the Swadesh 100 and 200 lists with additional words relevant to

comparative study of languages in Vietnam and Cambodia (Mann 2004). The 281-item list was expanded into the SIL Mainland Southeast Asia (MSEA) 436-item wordlist, with additional words relevant to Thailand and Myanmar. The expanded 462-item list was developed in 2008 by the Myanmar survey team. It removes items not native to Myanmar and adds in items from Matisoff's (1978) list of Tibeto-Burman core vocabulary (CALMSEA) and other items of local relevance to Myanmar. For research in Kachin and Shan States items not relevant to the area have been eliminated such words include those are not native to the people, or whose meanings are not clear in Burmese and the result is a 454-item wordlist.

2.1.2 Lexicostatistics

Lexicostatistics is a quantitative method used to measure the degree of similarity between two or more languages through comparison of their common vocabularies. Put briefly, a word in one variety is considered to be lexically similar to a word (with the same meaning) in another variety if they share 'enough' phonetically similar segments. Thus, lexical similarity is based on the cumulative similarity of phonetic segments. This similarity of phonetic segments is an approximation of word forms having descended from a common ancestor, which is they are cognates. The number of assumed cognate forms indicates the lexical similarity- expressed as the percentage of the total number of words compared- which is taken as a measure of the closeness of the languages (Fox 1995:279-291).

The lexico-statistical analysis in this thesis is adapted from Blair (1990). The first step is to identify the phonetic segments of each variety to be compared and specify which phones are considered phonetically similar to each individual phone. The process of determining phonetically similar segments is described and illustrated by Burquest (2001:41). Comparing words from different varieties is more complicated than a one-to-one comparison of individual segments, because sometimes language change involves the loss of segments such as the second element of an initial cluster, or a replacement of a final consonant by a suprasegmental feature such as tone. Blair proposed a two stage approach for handling such complexity. The first stage involves the categorization of each pair of phones and is laid out in Table 2. Note that the details of the categorization need to be adapted to accommodate the particular features of the speech varieties being compared.

Table 2: Criteria for comparing phone pairs (Blair 1990:31-32)

Category A: (a) Identical consonants

(b) Identical vowels or phonetically-similar (connected) vowels

(c) Phonetically-similar (connected) consonants that appear in a total of 3 or more word pairs (over the whole wordlist)

Category B: (a) Phonetically-similar (connected) consonants in fewer than 3 word pairs

(b) Vowels that are not connected

(c) [r]/[l]/[x] and nothing after another consonant

Category C: (a) Non-phonetically-similar (not connected) consonants

(b) A correspondence with nothing in fewer than 3 word pairs (over the whole wordlist)

Ignore: (a) The vowel [ə] between consonants

(b) A correspondence of a consonant or a vowel with nothing in 3 or more word pairs (over the whole wordlist)

(c) A correspondence between [ʔ]/ [h] and nothing for final consonants

(d) Suprasegmentals such as tones, breathiness

The second stage defines which combinations of categories for phone pairs are acceptable for the word pair to be considered lexically similar. These combinations are laid out in Table 3. The combination of categories for a particular word pair must match one of the specifications listed in Table 3 in order to be considered lexically similar.

Table 3: Acceptable category combinations for lexical similarity

No. of Phones		Category		
		A	B	C
1	=	1	0	0
2	=	2	0	0
2	=	1	1	0
3	=	3	0	0
3	=	2	1	0
4	=	4	0	0
4	=	3	1	0
4	=	3	0	1
4	=	2	2	0

No. of Phones		Category		
		A	B	C
6	=	6	0	0
6	=	5	1	0
6	=	5	0	1
6	=	4	2	0
6	=	4	1	1
6	=	3	3	0
6	=	3	2	1
7	=	7	0	0
7	=	6	1	0

No. of Phones		Category		
		A	B	C
4	=	2	1	1
5	=	5	0	0
5	=	4	1	0
5	=	4	0	1
5	=	3	2	0
5	=	3	1	1

No. of Phones		Category		
		A	B	C
7	=	6	0	1
7	=	5	2	0
7	=	5	1	1
7	=	4	3	0
7	=	4	2	1

Romaine (1994:5) suggested ranges for interpreting lexical similarity percentages as shown in Table 4.

Table 4: Guidelines for interpreting lexical similarity percentages (Romaine 1994:5)

Percentage range	Interpretation
Between 81% and 100%	Varieties both belong to the same language
Between 21% and 80%	Varieties both belong to the same language family
Between 0% and 20%	Varieties are from different language family

Blair (1990:23) states that if the results of a word list comparison show greater than sixty percent lexical similarity between two speech varieties, dialect intelligibility testing must be done. Blair described three possible situations as shown in Table 5.

Table 5: Typology of situations based on intelligibility and lexical similarity (Blair 1990:23)

		Lexical similarity	
		Above 60%	Below 60%
Inherent Intelligibility	Above 80%	Several very similar speech varieties may be referred to as similar dialects if inherent intelligibility is high.	Several dissimilar or slightly similar speech varieties may be referred to as different languages. (No dialect intelligibility testing is required.)
	Below 80%	Several very similar speech varieties may be referred to either as dissimilar dialects or different languages if inherent intelligibility is low.	

If word lists show less than sixty percent lexical similarity, then the speech varieties are referred to as 'different languages'. As a rule of thumb, no dialect intelligibility testing needs to be done between languages which have less than 60% lexical similarity.

If the lexical similarity is greater than 60%, then intelligibility testing is carried out. If the intelligibility test reveals less than 80% inherent intelligibility then the speech varieties are referred to as either 'dissimilar dialects' or 'different languages'. If the intelligibility testing shows more than 80% inherent intelligibility, then the speech varieties may be referred to as 'similar dialects'.

2.2 Intelligibility testing

Word lists and dialect intelligibility are used together to distinguish different dialect areas. One technique provides something the other lacks. Word lists provide information about the linguistic relationship between speech varieties. However, lexical similarity is limited when it comes to predicting intelligibility because it is based only on lexical analysis and cannot take into account syntactic features. Inherent intelligibility is the degree of understanding which speakers of one dialect have of a similar dialect because two dialects spring from the same linguistic stock, not acquired by exposure to it (Blair 1990:24). Comprehension testing or intelligibility testing is based on longer utterances such as sentences and texts and it helps delineate the existing intelligibility networks. Both are necessary for clear understanding of the situation in the region being surveyed (Blair 1990:23).

Intelligibilty between dialects is measured by a recorded text test (RTT). The test consists of a short text spoken by a mother tongue speaker of the language being tested. A subject from the other diaect listens to the text one time. The subject then hears the text a second time, with questions about the text interspersed in appropriate places throughout the text (Blair 1990:73). An intelligibility survey consists of four steps: (1) planning the survey (2) collecting the texts, (3) preparing test tapes, and (4) administering the tests (Simons 1983:5). Test scores can be evaluated by the relationship between the average score and standard deviation of test scores as shown in Table 6, which is an adaptation from Blair (1990:25).

14

Table 6: Interpretation of standard deviation (Nahhas 2007:70)

		Standard Deviation	
		High (greater than 12 - 15)	Low (less than 10 - 12)
Mean (Average Score)	High (above 80)	Situation 1 Many people understand the story, but not all.	Situation 2 Most people understand the story.
	Low (below 60)	Situation 3 Many people do not understand the story, some score rather higher than others.	Situation 4 Few people, if any, are able to understand the story.

High average RTT percentages with low standard deviations are taken to indicate that almost all the subjects adequately comprehend the variety represented by the recording. Low average RTT percentages are interpreted to indicate inadequate comprehension. If the RTT percentages are between 60% and 80%, it is not clear how well the subjects understand the variety being tested. If the average score is high and the standard deviation is high, it may indicate that some subjects have extensive contact with the tested speech variety, while others have little. Thus, those with low contact may not be able to understand that variety very well. It should also be noted that RTT measures comprehension of simple narrative texts and is only an approximation of how well subjects would understand more complicated texts or other genres.

Lexical similarity and dialect intelligibility are not always correlated. Joseph and Babara Grimes (1983) state that the lexical similarity of between two or more languages do not assure, that they will be mutually understandable to one another. However, Joseph Grimes maintains that lexically similarity is an indicator to measure the needs for a language development project. This is an intial step to analyse speech varieties that can identify these varieties that are sufficiently different as to require separate language development programs. Any speech varieties showing lexical similarity of less than 60% are assumed to need separate language development programs whereas languages with at least 61% lexical similarity should be further investigated by intelligibility testing.

Even when languages share high lexical similarity, intelligibility can be hindered by differences in high-frequency words such as grammatical particles. Such words are often not included in wordlists because lexical items such as nouns or verbs are easier to elicit accurately. Hanna (2010:1) discussed intelligibility between Central Thai and Tai Lue. The two languages are from the same branch of the Tai language family and they share many words that are the same, however, speakers of these two languages cannot understand each other. The major obstacle in communication is the differences in functor words. Hanna described twenty functional areas which proved to be the major obstacle for Thai and Tai Lue speakers to understand each other. The methodology of word list collection and lexicostatistic procedures used in this study will be described in section 3.3.1.

2.3 Phonological comparison

Phonological segments of two speech varieties are compared to show how related those varieties are. Comparisons can be done informally or using various quantitative methods. Simons (1983:67-69) reviews several phonostatistics methods which seek to quantify the phonological differences between speech varieties. He describes 12 phonostatistics methods and compares their various strengths and weakness, discussing their advantages over lexicostatistic methods. Duong (2003) also demonstrated an alternative approach of reconstructing the proto form and a quantitative method based on comparing the phonological innovations of various varieties. This thesis will use an informal or non-quantative comparison of the consonant inventories of Meung Yum with the inventories of Proto Wa (Diffloth 1979) and Standard Wa (Watkins 2002).

2.4 Sociolinguistics

Sociolinguistics is the study of language and society in order to understand how languages function in communication (Wardhaugh 1998:12). In this thesis, several topics are investigated including bilingualism, language vitality, language choice, language attitudes, contact with insiders and outsiders, literacy rate, interest and attitude in vernacular language development and identifying the prestige dialect. These are described in the following subsections.

2.4.1 Bilingualism

The term 'bilingual' is used not only to mean the ability to speak two different languages, but broadly used to cover multilingual situations where individuals have some competence in three or more languages. Spolsky (1998:45) defines a bilingual person as 'a person who has functional ability in a second language'.

Blair (1990:52-53) wrote that bilingualism is not uniformly distributed in a community. Individuals and sections of any community could be bilingual to different degrees. Factors which influence bilingualism include people's motivation and the amount of contact they have with speakers of the second language. Various social characteristics often correlate with amount of contact such as age, sex, education, and frequency of contact.

Blair (1990:51-65) describes several methods for surveying bilingual ability of an individual and also lists advantages and disadvantages. Self-evaluation questionnaires consist of a series of questions asking each person whether or not they are able to perform a particular task using the speech variety of interest. The questions are usually asked in order of increasing difficulty, that is, the later in the sequence the question appears, the greater the command of the second language required to carry out the task described in the question. In this thesis, a set of seven questions are used which give insights into the communication proficiency of Meung Yum speakers in various LWCs.

2.4.2 Language choice

Fasold (1984:180-181) discussed three kinds of language choices: code switching from one language to another language; code-mixing or borrowing which is the use of pieces of one language while a speaker is basically using another language; and variation such as accents within the same language. Language choices can happen for monolingual and bilingual speakers. According to Fasold, these three kinds of language choice occur as a continuum and they cannot be separated from one another.

Fishman (1964) used a sociology approach. He proposed that a *domain* is an institutionalized context in which one speech variety is more likely to be chosen than another. Domains are defined by various factors such as location, topic and participants. If an individual is at home talking to another member of their family

about an everyday topic then that individual could be said to be in their 'family' domain. Evaluating the domains in which people choose difficult languages is a relatively objective way to measure language choice.

2.4.3 Language vitality

Language vitality is measured by the situation of how much people use the language in communication. Languages that continue to be actively used are said to be 'alive' in contrast with dead languages. A language is dead when the speakers of it have either all died out or ceased to use the language (Wardhaugh 2002: 37). Language maintenance, shift and death are three terms to describe stages of language vitality (Larson 2002). Language maintenance is when the community collectively decides to use the language(s) it has traditionally used. Fasold (1984: 213) states that language shift is the situation in which 'a community gives up a language completely in favour of another one'. The ultimate result of the process of language shift is language death.

Landweer (2002:20) proposed eight indicators of ethnolinguistic vitality through the experience of SIL in nearly 300 languages in Papua New Guinea. These are listed in Table 7.

Table 7: Ethnolinguistic vitality indicators (Landweer 2002:20)

No.	Description of ethnolinguistic vitality indicator
1	Relative position on the urban-rural continuum
2	Domains in which the language is used
3	Frequency and type of code switching
4	Population and group dynamics
5	Distribution of speakers within their own social networks
6	Social outlook regarding and within the speech community
7	Language prestige
8	Access to a stable and acceptable economic base

Edwards (1997:34) pointed out that home is the most important domain in the maintenance of a minority language. If the minority language is not spoken at home, it is unlikely to be spoken by next generation. This will end with the entire people ceasing to use their language in favor of the majority/dominant language, and then language death will result. He also makes the distinction between domains of

18

necessity (such as home, school and the workplace) which typically relate to the central aspects of peoples' lives and domains in which a person's participation is more voluntary or sporadic.

Fishman (1991) suggested the degree of language vitality can be evaluated by several factors: (1) intergenerational language transmission; (2) absolute number of speakers; (3) proportion of speakers within the total population; (4) trends in existing language domains; (5) response to new domains and media; and (6) materials for language education and literacy. In this thesis, intergenerational transmission was investigated by asking about whether the language is being passed on children by their parents and whether children use the language when playing together. Trends in existing domains were investigated by asking subjects to predict whether children of the future would still be speaking Meung Yum.

2.4.4 Language attitudes

Language attitudes are the feelings people have about their own language or the languages of others (Crystal 1992). Attitudes of a person to his/her own variety can effect how much he/she uses it in communication. (Fasold 1984:147-152) describes various methods for language attitude assessments. Using a questionnaire is a common method and direct approach. It means asking subjects how they feel towards a particular speech variety. This method is simple but the validity of the responses is questionable. An indirect method which still uses a questionnaire, is asking a series of questions that relate to language attitudes, more indirectly. Blair (1990:113) suggests two kinds of language attitude questions. One kind of question needs the person to give the name of a language as an answer and the other kind needs a 'Yes' or 'No' response. Data generated by these kinds of questions result in a good indication of either a positive or negative attitude toward a speech variety. This is a method used in this study.