Chapter 3
Theoretical Foundations

This chapter provides theoretical frameworks for the methodologies used in this thesis. Section 3.1 gives a brief description of lexical comparison. Section 3.2 describes intelligibility testing as measured by Recorded Text Testing. Section 3.3 discusses sociolinguistics, focussing on bilingualism, language choice, language vitality and language attitudes.

3.1 Lexical Comparison

This section explains about wordlists and the lexical similarity method used in this study.

3.1.1 Wordlists

Wordlists are used to elicit the words of a language and compare speech varieties in a language survey. The wordlist is determined in advance, can be designed to be specific for the language family of the varieties being surveyed and allows the researcher to collect a certain amount of language data in a limited time. The researcher transcribes and records the words for future study. Wordlists are helpful for discovering lexical similarity and describing phonology. However, wordlist does not measure intelligibility.

Swadesh (1955) provided a 100-wordlist that should be suitable for all languages. It comes up with a list of basic or core words like body-part names, kinship terms, natural objects, simple action verbs, lower numerals, etc (Swadesh 1952:122-126; 1955). Mann (2004) examined wordlists used in Southeast Asia and lists culturally relevant to the language families of Mainland Southeast Asia. A total of 511 different lexical items were identified among the various lists. Mann gave each item a ranking based on how many different lists it was included in. Mann proposed that items in the highest rank be given priority when comparing languages of the region.

The wordlist used in this study is the Myanmar Tibeto-Burman wordlist. This wordlist is derived from the SIL MSEA 281-item wordlist, which was based on the

Swadesh 100 and 200 lists with added words for comparative study of languages in Vietnam and Cambodia (Mann 2004). The SIL MSEA 436 wordlist was the expansion of the SIL MSEA 281-item wordlist with added words relevant to Thailand and Myanmar. The Myanmar Tibeto-Burman wordlist was developed in 2008 by the Myanmar survey team.

3.1.2 Lexicostatistics

Lexicostatistics is a method developed mainly by Morris Swadesh. It is an attempt to figure out whether two or more languages are genetically related and how to group any known language varieties. It measures the degree of similarlity between languages or language varieties by comparing common items in wordlists.

Bennett (1998:34) proposes that lexicostatistics is useful for comparing the degree of liguistic relationship between languages. Brown (2008:248) asserts that lexicostatistics counts lexical similarity percentage for common roots between two languages by using core or basic words. Fox (1995:279-291) claims that lexical similarity is measured by the cumulative similarlity of phonetic segments. This similarlity is an estimation of word forms that are genetically related. This gives a measure of the closeness of the language varieties.

The purpose of using lexicostatistics in this research is to determine the degree of linguistic similarlity among the selected Bisoid varieties and to generalize the level of intelligibility which exists among the varieties. As Blair states, Word lists and provide information about the linguistic relationships between the speech varieties insofar as these relationships are not blurred by borrowing (Blair 1990:23).

The method of lexical analysis in this study is adapted from Blair (1990:29-33). First, phonetically similar segments of each language variety are individually identified, examined and chosen to be compared. Burquest (2004:41) describes the steps how to examine phonetically similar segments in the context of phonemic analysis. Langauge change includes the loss of segments. Blair proposes how to handle such problems. Table 7 presents ranges for interpreting lexical similarity percentages (Romaine 1994:5).

Table 7 Guidelines for interpreting lexical similarity percentages (Romaine 1994:5)

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

Joseph Grimes (1988:33) suggests that there may need separate language development programs if speech varieties that show lexical similarity of less than 60%. However, intelligibility testing should be conducted if the lexical similarlity is above 60%. Blair (1990:23) also recommends that dialect intelligibility should be undertaken if the results of a word list comparison show higher than 60% lexical similarity between two speech varieties. Ranges for interpreting lexical similarity percentages suggested by Blair are given in Table 8.

Table 8 Percentage of similar Words (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.	

Speech varieties are assumed as different languages if the wordlists show less than 60% lexical similarity. There is no need for dialect intelligibility testing in this case. However, researchers should carry out intelligibility testing if the result of lexical similarity percentage is higher than 60%. The speech varieties are regarded as different languages if the result of intelligibility testing shows less than 80% of inherent intelligibility. The varieties will be regarded as similar dialects if the intelligibility testing results reveals more than 80% inherent intelligibility.

3.2 Intelligibility Testing

Blair (1990:23, 24) explains that word lists and dialect intelligibility testing are used together to distinguish different dialect areas. One technique complements the other. Word lists examine the linguistic relationship between speech varieties to the degree that these relationships are not blurred by borrowing. However, when it comes to predicting intelligibility measuring lexical similarity is limited. It is because the lexical analysis is focused on words (usually nouns, verbs and adjectives) and cannot take into consideration grammatical function words or discourse features.

To test the level of comprehension between speech varieties or dialects, a Recorded Text Test (RTT) is used. In the RTT subjects listen to a recorded story from another variety and answer comprehension questions about the story in order to assess how well they understand that variety (Blair 1990:73). Based on the evaluation of the subject's answers to comprehension questions in their own language, the subject's general comprehension of the text can be inferred. Simons (1983:5) explains the four steps of an intelligibility survey: (1) planning the survey, (2) collecting the texts, (3) preparing test tapes, and (4) administering the tests.

The development and administration of the RTT in this thesis basically followed as described in Blair 1990:73-85; Nahhas 2006, 2007. The test consists of listening to a story, told by a native speaker, with comprehension questions, asking in a language of wider communication (LWC) and translating with the help of a translator. Table 9 gives the relationship between the average score and standard deviation of test scores, which can be evaluated by test scores (Blair 1990:25).

Table 9 Interpretation of RTT results (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 indicate that almost all the subjects comprehend the recorded variety. Low average RTT percentages are interpreted to indicate that many people do not understand the recorded story. If the average score is high but the standard deviation is also high, it may indicate that some subjects have extensive contact with the tested speech variety, while others have little. Therefore, those with low contact are not able to understand that story well. RTT is limited in that it normally tests comprehension of simple narrative texts. Thus, it is not necessarily a good indication of how well subjects would comprehend more complicated texts.

Lexical similarity and dialect intelligibility are sometimes disconnected. Joseph and Babra Grimes (1983) state that high lexical similarity between two or more speech varieties does not assure that they are mutually intelligible to one another. And yet, Joseph Grimes insists that lexical similarity can have a part to play in identifying projects that need a language development project. It can analyze speech varieties and indicates whether the varieties need separate language development programs. If the lexical similarity of the speech varieties is less than 60%, it is assumed that the varieties need separate language development programs. If the similarity is at least 61%, then intelligibility testing should be followed.

Even when the lexical similarity of speech varieties is high, discourse features can hinder the intelligibility. Hanna (2010:1) talks about 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. However, functor words hinder speakers of the two languages from understanding. Hanna presents twenty functional areas which hinder the speakers from comprehending.

3.3 Sociolinguistics

Sciolinguistics is the study of of relationships between language and society with the purpose of understanding of language structure and of how languages function in communication (Wardhaugh 2006:13). The following subsections briefly define bilingualism, language choice, language vitality and language attitudes.

3.3.1 Bilingualism

Crystal (2003:51) suggests that the term 'bilingual' refers to 'a person who can speak two languages'. Studies of bilingualism involve the analysis of social, psychological and national concerns such as the social status of different languages. Crystal

mentions two types of bilingualism. In the first type of bilingualism, a major group learns a second language without threatening the first language. In the second type of bilingualism, the first language is replaced by the second language.

Blair (1990:51) suggests that bilingualism is distributed in a complex way throughout a community. The degree of bilingualism will most likely be different in subgroups with different social status. Blair (1990:54-64) discusses various factors that have been known to correlate with bilingualism. Surveyors should consider these various factors when they are investigating bilingualism in a community.

Blair (1990:67-106) discusses some methods for surveying individual bilingual ability and suggests their strengths and weaknesses. Self-evaluation quesionnaires include a series of questions which ask each subject whether they can perform particular tasks using the speech variety of interest. The tasks earlier in the sequence (e.g. buying something) are simpler than tasks later in the sequence (e.g. describing a conversation overheard in the market).

3.3.2 Language choice

There are choices in using language. Language choice is a concept that is basically related to the language attitudes of individuals or a group of speakers. Fasold (1984:180-181) presented three kinds of language choices: code switching; code mixing and variation within the same language. Code switching involves a choice between two languages. Code mixing involves the use of pieces of words or phrases. Variation within the same language includes variations of pronunciation, vocabulary and grammar. Speakers will choose them depending on the particular social setting.

Researchers have studied the issue of language choice in different approaches. As language choice is constrained by sociological factors, sociologists find ways to explain choice in terms of abstract social constructs. Anthropologists find ways to identify 'the values of a sociocultural group and the cultural rules of behavior that reveal those values' (Fasold 1984:192).

Fishman (1964) proposed a sociological concept of a 'domain'. He said that a domain is 'an institutionalized context in which one speech variety is more likely to be chosen than another' (1964:32-70). Domains are defined by various factors such as topic, location and participants. An individual could be said to be in their 'family' domain if they are at home talking to another member of their family about some everyday topic.

3.3.3 Language vitality

Language vitality means the continuing use of a language. Edwards (1985:49) suggests that language vitality is a measurement of language ability. Language serves the needs of a community. Language satisfies the needs of a society. The vitality of the first language is becoming weak when a language of wider communication is widely used in the community. If individuals in the community use their monther-tongue at home and in public meetings, the vitality is strong. If they do not use their language at home and in public meeting, the vitality is weak.

The continuing use of a language is thus directly related to its benefit to that community. If speakers of a language variety do not use that language anymore, it will become a dead language. Researchers have classified language vitality into three categories: language maintenance, language shift and language death. Language shift happens when a community gives up speaking their language and replaces it by another language (Fasold 1984:213). This language shift takes place over time. Language maintenance means that a speech community chooses to continue using their language.

Table 10 presents eight ethnolinguistic vitality indicators that have been 'discovered, developed and documented in the Papua New Guinea context' (Landweer 2002:20).

Table 10 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 (1995:34) said that the home domain is probably the most important domain for all languages in determining whether the language continues to be maintained, although there may be other domains which are necessary for some languages. If a second language variety replaces the first language in certain domains, language shift is taking place. If the second language dominates all domains, language death will take place.

3.3.4 Language attitudes

Preservation of a language variety is very much dependent on language attitudes. If speakers of a particular variety feel that their variety is less prestigious than a second language, the maintenance of their variety is threatened. Crystal (2003:256) defines language attitudes to be 'feelings' of people towards their own langauge variety. It is thus important to know whether a community values their language variety or not because it helps evalute the need for a language development program.

Fasold (1984:34-60) defines language attitudes to be 'the ways in which a speech comminity perceive the relevance and status of their language, often reflecting their attitudes about themselves relative to other groups'. Language attitudes determine language maintenannce, language shift and language death.

Blair (1990:107-109) suggests that the language attitudes of a community of speakers towards their mother tongue vary from one another. This is true also for related varieties and for unrelated varieties which might be used as languages of wider communication in the area.

Although language attitudes may range over a continuum they are often categorized using a scale like the seven-point scale given in Table 11.

Table 11 Levels of language attitudes (Blair 1990:109)

Strongly positive	Positive	Midly positive	Neutral	Mildly negative	Negative	Strongly negative

Fasold (1984:149-152) presents different ways of investigating language attitudes. One approach is conducting a questionnaire. This method inquires about subjects' attitudes towards a speech variety. The attitudes can be indirectly interpreted according to the subjects' answers. These answers infer their positive or negative attitudes towards a particular speech variety.

Having presented the theoretical foundations for the methodologies used in this study, the designs for the surveys on which this thesis is based will be presented in Chapter 4 and Chapter 6.