

CHAPTER 3

THEORETICAL FRAMEWORK

3.0 Introduction

This section covers the theoretical framework applied in this thesis. It includes two main areas: lexicostatistics and sociolinguistics including a social network analysis.

3.1 Lexicostatistics

Lexicostatistics is a quantitative method used to define the degree of similarity between languages. This technique applies statistics to linguistics. The goal of lexicostatistics is to establish linguistic relationships on the basis of a quantitative comparison of vocabularies. The closeness or degree of similarity of languages is shown through the comparison of vocabulary. The more shared vocabulary, the closer the relationship between them is likely to be.

Vocabulary is considered to be similar if a pair of words shares the same or similar consonants and vowels. Thus, this similarity is based on phonological features. This similarity of phonological features is an approximation of forms having descended from a common ancestor or cognates. The number of assumed cognate forms indicates the lexical similarity showing the closeness of the languages (Fox 1995). Similarly, Busenitz and Martens (1979) described lexicostatistics as follows:

Lexicostatistic methods start with a basic word list, such as the Swadesh word list. The survey team elicits the words on the list in each test location, and then compares how similar the vocabulary items are throughout the area. The term lexicostatistics is usually a synonym of glottochronology, that is, a method used to compute

time depth. But here it is used in the broader sense of simply comparing lexical items (1979:11) .

Romaine (1994:5) stated that, in general, this method regards a dialect within a language as sharing between 81 to 100 percent cognates. If there are between 28 and 81 percent cognates, then the varieties are usually considered as languages within a family. The greater the number of cognates, the closer the relationship.

3.1.1 Lexicostatistic methodology

Sanders (1980:31) recommends a format for collecting wordlists that makes it easy for the researcher to include several languages for comparison. The format is shown in Figure 20.

	Word ₁	Word ₂	Word ₃	Word ₄	Word ₅
Language ₁					
Language ₂					
Language ₃					

Figure 20: Format for wordlist comparison (Sanders 1980:31)

When more than one form is given, Sanders suggests alternative solutions: put down the first word given, write all the words, or get a consensus opinion from the group.

Blair (1990:31-33) describes a lexicostatistic methodology based on criteria for determining lexical similarity. Words are compared between related languages. These words are compared on a phone-by-phone basis to determine which criteria applies to the phone pair. Modifying the criteria from Blair slightly for features of Asian languages, Mann (2001) proposes the following criteria:

- Category 1: (a) Exact consonant matches,
 (b) Vowels or diphthongs differing by 1 or fewer features,
 (c) Phonetically similar consonants in 3 or more word pairs, and
 (d) A deletion in three or more word pairs.
- Category 2: (a) Phonetically similar consonants in less than 3 word pairs, and
 (b) Vowels differing by 2 or more features.
- Category 3: (a) Non phonetically similar consonants, and
 (b) A correspondence with nothing in less than 3 word pairs.
- Ignore (a) A regularly occurring epenthesis.

Figure 21: The criteria for lexical comparison (Mann 2001)

After the word pairs are compared, the criteria are added up to determine lexical similarity. Next, the results are compared with a phone table which establishes the minimum conditions pairs of words must satisfy in order to be considered lexically similar. The phone table is shown in Figure 22. Word length in the first column is number of phones.

Word Length		Category 1	Category 2	Category 3
2	=	2	0	0
3	=	2	1	0
4	=	2	1	1
5	=	3	1	1
6	=	3	2	1
7	=	4	2	1
8	=	4	2	2
9	=	5	2	2
10	=	5	3	2
11	=	6	3	2
12	=	6	3	3

Figure 22: Word Length and Linguistic Similarity (Blair 1990:32)

3.1.2 The wordlist

Lehmann (1962:107-108) indicates that lexicostatistics may be used to study language loss, specifically, the loss of vocabulary items. He also suggests that basic vocabulary should be used. Basic vocabulary includes lower numerals, pronouns, body parts, animals, plants, heavenly bodies, etc.

Sanders (1980:26-27) suggested that some words on the wordlist should be eliminated. For example, cases where there are no vernacular terms such as 'snow' or 'horn'. Items which may be expressed as a phrase for items already elicited, or items borrowed from another language, should be ignored. Each wordlist item should be clearly defined, for instance, 'neck' should be specified as either 'throat' or 'nape'.

Rensch (1992:13) explains the kinds of inconsistencies that frequently emerge in wordlist collection. One of them is that dissimilar words are given when similar words exist. This inconsistency can lead to five to ten percent of the items on wordlists from any pair of related dialects being unnecessarily different. This inconsistency can happen for the following reasons: a generic word is given instead of the specific word, the words may be synonyms or the person who gives the wordlist may misunderstand the words used in elicitation.

Vocabulary can be replaced by borrowed words. If one of the words used for comparison is borrowed from Chinese, or any LWC (language of wider communication), the results of the comparison may not give an accurate representation of the lexical relationship between the languages under investigation (Fox 1995). The researcher should also work to minimize the effects of interference from the language of elicitation.

3.1.3 Patterns in lexicostatistics

Simons (1980:107-134) describes how patterns of language convergence and divergence can be recognized in a matrix of lexicostatistic relations. Convergence

refers to the process by which two speech varieties become more like each other, while divergence is the process by which they become less like each other. For instance, if language B is 50% cognate with A and C, but A is only 40% cognate with C as shown in Figure 23 then the following diagram can be used to depict the situation.

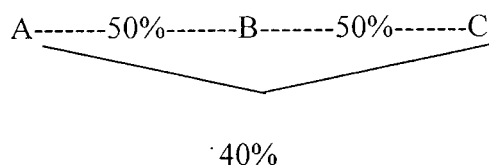
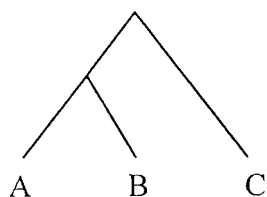


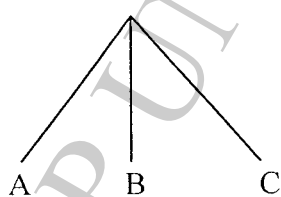
Figure 23: Cognate Percentages (Simons 1980:108)

Based on this situation, three theoretical patterns of divergence are possible as shown in the Figure 24.

Tree diagram based on
on cognates with A:



Tree diagram based on
cognates with B:



Tree diagram based
cognates with C:

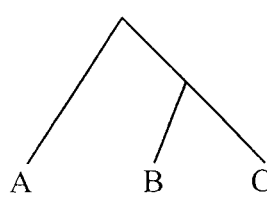


Figure 24: Tree diagrams (Simons 1980:108)

Time depth can be used to indicate when speech varieties divided either in absolute time or relative to each other. This analysis only considers when varieties separated related to one another. The divergence pattern can be used to infer the time depth as shown in Figure 25. Note that the scale shows relative time depth with 1 being more recent splits and 3 being more distant.

Time depth

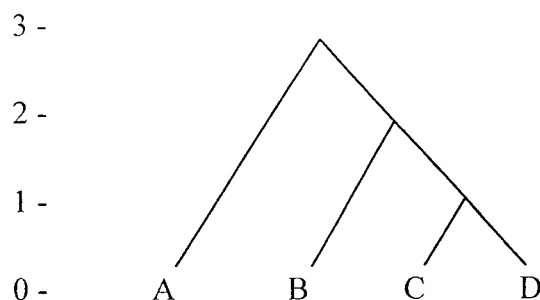


Figure 25: Divergence and time depth tree diagram (Simons 1980:110)

The same relations are shown in the matrixes in Figure 26, where the number in each cell represents the time depth.

A	B	C	D
3			
3	2		
3	2	1	

A	B	C	D
73			
73	81		
73	81	90	

Figure 26: Matrixes of time depth as cognate percent (Simons 1980:110)

Generally, it is assumed that cognate percentages will be lower with greater time depth³. The deeper the time depth, the lower that percentage is likely to be and the longer the column of equal percentages will likely be. Simons (1980:110) suggests patterns of divergence from the matrix of lexicostatistic relations. These patterns

³ It must be noted however that lexical change is affected by language contact, and processes such as lexical diffusion, borrowing, and semantic shift are closely related to language contact. This contact may not be uniform over time.

are depicted in Figure 23. The first pattern containing three diagrams shown in Figure 27 is the basic convergence pattern.

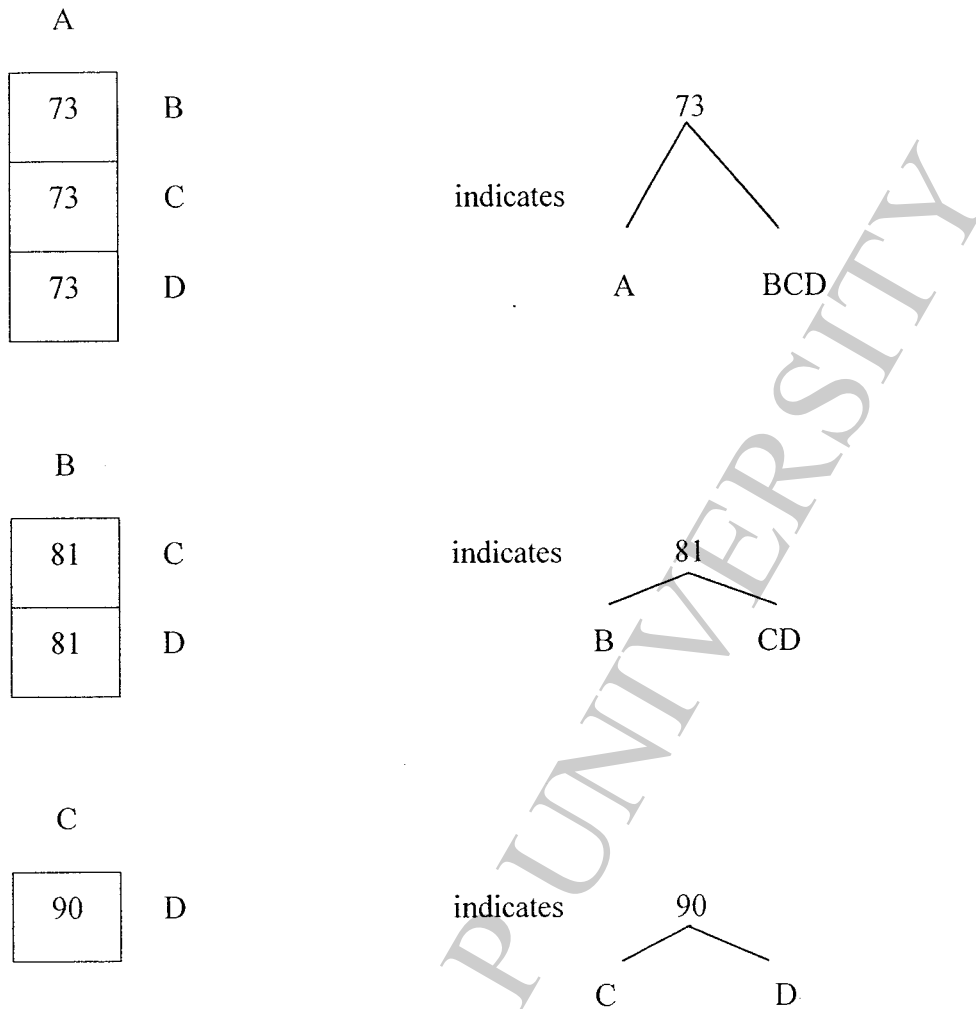


Figure 27: Divergence tree diagrams (Simons 1980:112)

The second pattern is a basic convergence pattern. This pattern occurs where there are two equal and higher percentages and one lower percentage between three speech varieties. This pattern is illustrated in Figure 28.

Time depth		Cognate percentage	
A		A	
1	B	90	B
2	1	81	90
	C		C

Figure 28: Convergence patterns (Simons 1980:114)

Based on the percentages shown in Figure 28 it is assumed that A and B split at time depth one as did B and C. The convergence tree diagram is shown in the Figure 29.

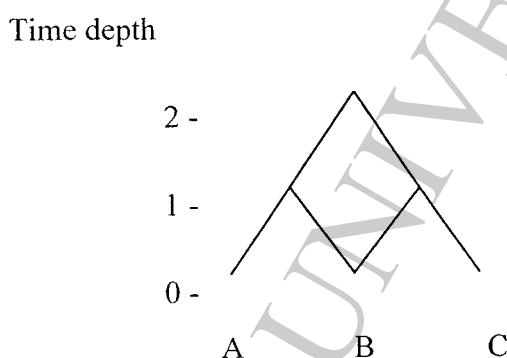


Figure 29: Convergence tree diagram (Simons 1980:114)

The pattern in Figure 29 could be explained historically in at least three ways.

- 1) A and B split from C, and then B split from A. Later B had contact with C, and therefore B converged with C and became more like A.
- 2) B and C split from A, and then B split from C and converged with A through contact.
- 3) A, B, and C split from one another at the same time, but B maintained contact with A and C, while A and C did not have contact. Thus, A and C have diverged.

Simons (1980:115-125) further explains that when there are more than three speech varieties, the basic convergence patterns cannot resolve such a case. He, therefore, introduces three more patterns of convergence called chaining, dominance and sporadic convergence.

Chaining typically occurs when speech varieties are located in close geographic proximity. Each dialect has contact with speech varieties on either side of the chain. This pattern shows the highest cognate percentages between neighboring speech varieties and the cognate percentages are lower and lower as the dialects are further and further apart along the chain⁴. Figure 30 depicts the geography of a chaining pattern.

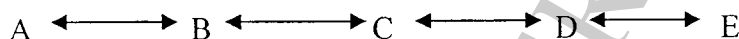


Figure 30: Geography of the chaining pattern (Simons 1980:116)

The chaining pattern is shown by three diagrams in Figure 31. The diagram on the left shows time depth; the middle diagram shows lexical percentages; the diagram on the right is a lexical tree diagram.

⁴ Thus, the lowest cognate percentages are between speech varieties on the extreme ends of the dialect chain.

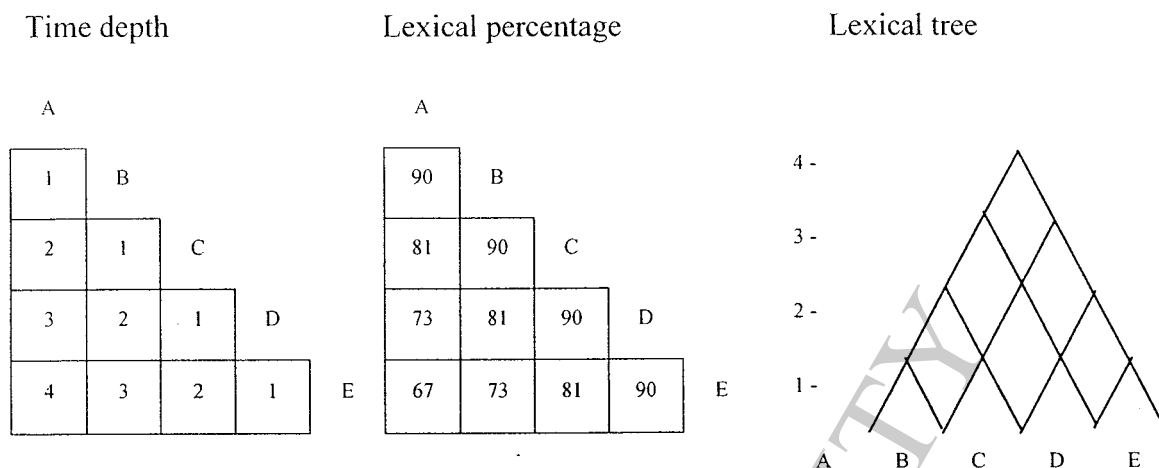


Figure 31: The chaining pattern (Simons 1980:116)

A dominance pattern typically occurs when one speech variety is geographically central to the other speech varieties. This pattern depicts the highest lexical percentages between the peripheral speech varieties and the central speech variety. Figure 32 shows the geography of the dominance pattern.

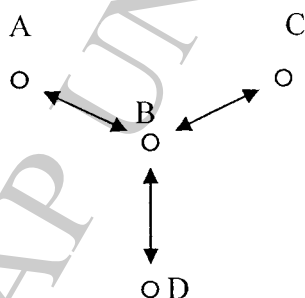


Figure 32: Geography of the dominance pattern (Simons 1980:117)

The dominance pattern is shown by two diagrams in Figure 33. The diagram on the top left shows time depth, the top right shows lexical percentages and the diagram on the bottom is a lexical tree.

Time depth

A			
1	B		
1	3	C	
1	3	2	D

Cognate percentage

A			
90	B		
90	73	C	
90	73	81	D

Lexical tree

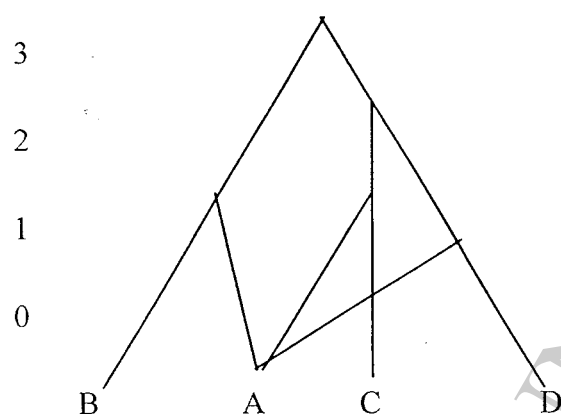


Figure 33: The dominance pattern (Simons 1980:117)

Sporadic convergence typically occurs when language contact through 'borrowing' takes place. The vocabulary between speech varieties that borrow and those speech varieties from which words are borrowed become more similar. If one language has a significantly higher lexicostatistic percentage, it shows a convergence relationship the language has. For instance, in Figure 34, there is a split between ABCD and EF. But the percentage between C and E is obviously higher than others. This indicates a potential convergence relationship between C and E.

A	90	B		
	81	81	C	
	73	73	73	D

	67	67	<u>76</u>	67	E	
	67	67	67	67	81	F

Figure 34: Sporadic convergence in a divergence pattern (Simons 1980:118)

3.1.4 Correlations between lexical similarity and intelligibility

Barbara Grimes (1984:17) claims that high lexical similarity does not always correlate with high intelligibility, although comprehension is always poor when lexical similarity is low. Intelligibility is complex, with many factors that influence intelligibility besides lexical similarity, such as differences in grammar, semantics, rate of speech, sociolinguistic factors, etc. Even when intelligibility tests are conducted, the results may not be generalizable to the target population. This is because bilingual proficiency within a community varies from one person to another. Each individual has different opportunities to learn, use and process a language. Thus, the sample must account for the variation in the community.

One intelligibility case study is cited in the Tenth Edition of the Ethnologue (B. Grimes 2004). This study gives the percentages of lexical similarity and intelligibility based on field studies. One of the results shows that with lexical similarity above 60%, intelligibility may range from 31% to 98%. This indicates that it is difficult to predict intelligibility based on lexical similarity. However the intelligibility of closely related speech varieties tends to correlate with lexical

similarity. Grimes (1998:32) suggests that where language development is being decided, considering lexical similarity is an effective strategy. He recommends that if lexical similarity is below 60%, there is a need for separate language development programs. Where lexical similarity is equal or greater than 60%, intelligibility testing should be conducted to determine if the speech varieties can share common literature.

3.2 Sociolinguistics

Wardhaugh (1998:10) defines sociolinguistics as the study of the relationship between language and society. Language and society are both directly and indirectly related. Since one of the goals in this survey is to investigate the uses of language among the Tai Nua, sociolinguistic investigation techniques are applied.

3.2.1 Language Choice

Language choice is possible when there is more than one language used in a community. Fasold (1984:180-181) proposed three kinds of choice: code-switching, code-mixing and borrowing.

According to Gal (1998:247), “code-switching is a conversational strategy used to establish, cross or destroy group boundaries; to create, evoke or change interpersonal relations with their rights and obligations.” Code-switching takes place when a person who can speak more than one language chooses to speak one of these languages in certain circumstances.

Fasold (1990:180) states that code-mixing takes place when a person uses a part of one language while he is using another language. Borrowing is when a person uses foreign words or phrases in the sentences or conversation.

According to Edwards (1995:72) “Language choice is non-random, and heavily influenced by external constraints.” He also showed factors influencing language choice in Paraguay where over 90% of the population is bilingual in Guarani, the traditional language, and Spanish. The situation Edwards describes is shown in

Figure 35.

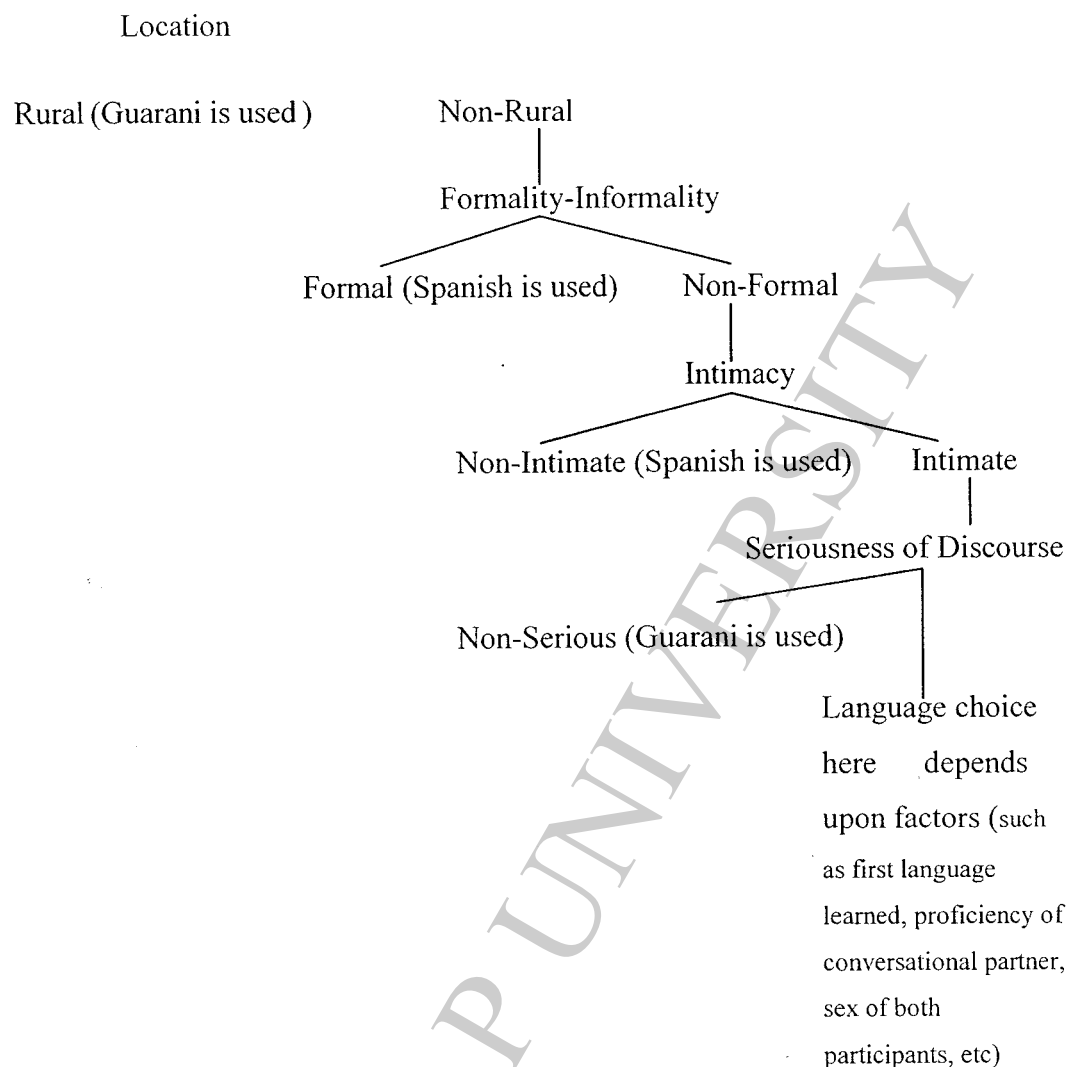


Figure 35: Factors influencing language choice in Paraguay (adapted from Edwards 1995:73)

Figure 35 shows that language choice in Paraguay depends on the situation of the speaker. For example if they are in a rural area, speakers prefer to use Guarani rather than Spanish. These speakers also consider the relationship between the speaker and the listener. If they have an intimate relationship, then Guarani is used. In the same way, if they have a discussion about something non-serious,

they will use Guarani while Guarani or Spanish might be used (depending on several factors) for discussion of a serious topic.

Fishman (1964:38) introduced the notion of domains in language choice. Each domain consists of different factors: location, topic and participants. The speaker will choose to use the most appropriate language for the domain.

Fasold (1984:186) said a typical domain is that of the family. He proposes a cline of ratings for 'mother tongue/first language' in descending order by domain. The domains are family, neighborhood, friend, transportation, government, employment, and education. Fasold states that the mother tongue is used most in the family domain and least in the domain of education.

3.2.2 Language Vitality

Larson (2002) states "the goal of language is communication." If people continue to use a language to communicate, it remains vital. Language vitality is, therefore, the extent to which a language continues to be used in a community. There are three terms that are used to describe the vitality of the language: language maintenance, language shift and language death.

Fasold (1984:213) defines language shift as taking place when a community gives up the use of one language in favor of another. It occurs when a new language is chosen for use instead of the old one. Language shift may result in language death. Language death occurs when a community shifts to a new language so that the old language is no longer used.

Language maintenance occurs when a language continues to be spoken as it had been traditionally used. Usually, when a community begins to speak a new language, it is used in certain domains instead of the old one. This is a sign that language shift is in progress. However, language maintenance often occurs in bilingual and multilingual communities as each language has its own domains. Larson (2002) reminds us that if the traditional language is still used at home and

in the community for communication, the language can remain vital, even if the new language or the language of wider communication is used in many other domains. If the traditional language is not used in the home, its vitality is weak. Another important indicator of language vitality is the language that the children and the young generation uses. If they use the traditional language, it means the language is still vital.

3.2.3 Language attitudes

Fasold (1984:145-147) states that language attitude is based on a mentalist or internal state. It means the attitude of a person is complicated and might not be obviously shown. For example if one was asked directly 'Do you like your language?', they might say 'Yes,' though they don't really care about it. But the real attitude could be found when the subject was asked something like 'What language do you use at home?' If a person's attitude is revealed, one would be able to make predictions about his behavior related to those attitudes with some degree of accuracy.

Blair (1990:109) states, "Language attitudes are the attitudes which a person holds towards the various speech varieties which are known to that individual." He presents a continuum from positive to negative, as shown in Figure 36,

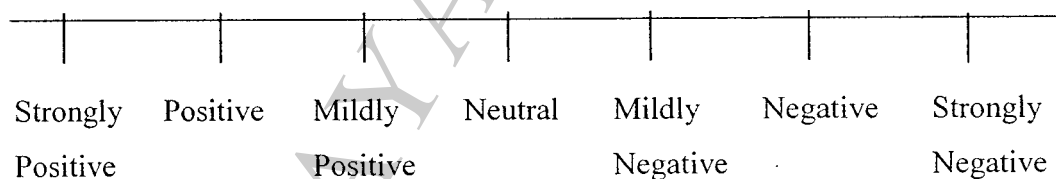


Figure 36: A continuum from positive to negative (Blair 1990:109)

One of the methods to determine language attitudes mentioned by Fasold (1984) is by administering 'questionnaires', which are explained more in the next chapter. Fasold (1984:145-147) also mentions 'interviews' which contain open-questions

and do not require the questionnaire and response to be written (as do questionnaires).

Discovering language attitudes is very helpful to aid in determining the acceptable language for an education program. The attitude a person has toward a language may influence the way he answers the questions. Sometimes a person will say he doesn't understand a language even when he does understand, because he has a negative attitude toward the language. But if he has a positive attitude toward the language, he is more likely to try to understand.

3.2.4 Social Network analysis concepts

A social network is seen as “ a web of cultivated and maintained relationship ties that spread through society, serving to link individuals to one another, forming a meaningful, purposeful community for these linked individuals” (Graham 2000:45)

Rueck (2005:1) describes that the social network has two basic elements. First, the actor or node is an individual person or other kinds of units in society such as households, clans or villages. Another one is links or relations which are between the actors. The arrowheads of the links indicate the direction of the actor, which can also go both ways. If the actor sends out the relations or links to other actors, it is called out-degree and if the links are sent to him, it is called in-degree. When there are many links between actors in both sending and receiving, there are a high number of connections within the network. The number of links reflect the density of the network. Figure 37 shows low and high density network.

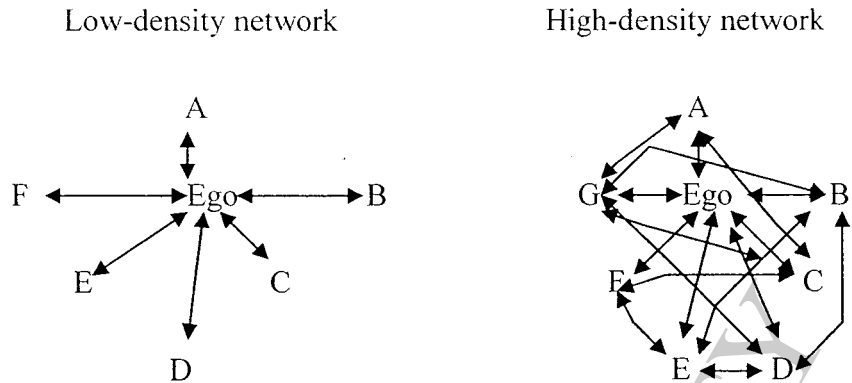


Figure 37: A low-density network and high-density network

Rueck (2005:4) explains ways to discover sub-groups of the social network. There are two approaches: top-down approaches and bottom-up approaches. One of the top-down approaches is called K-cores. It is “a maximal sub-group in which each point is adjacent to at least k other points: all the points within the k -core have a degree greater than or equal to k ”. Thus, if it is 2-core, it means each actor has a direct link to at least 2 other actors in the core. As for the bottom-up approaches, it starts from an actor and seeks to discover what nested groups he is part of. The term ‘clique’ is a group of individuals that have connections in every possible pair of points.

The term centrality is seen when an actor has a higher degree of connections than other actors. The most centralized network is known as a star network.

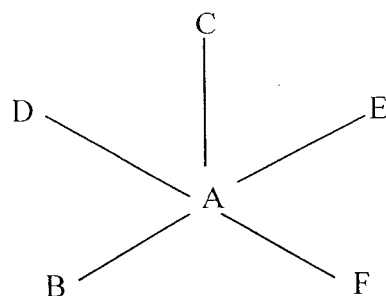


Figure 38: A star network

The most central actor has a shortest distance and is able to pass a message to other actors in the network quickly.

3.2.5 Bilingualism

Spolsky (1998:45) states that a bilingual person “is a person who has some functional ability in a second language.” However, one could be very limited in one or more domains or very strong in both languages and still fulfill the conditions of this definition.

Blair (1990:51) says, “the goal of a study of community bilingualism is to find out how bilingual the population of a community is.” He also presents “bilingualism evaluation methods.” One of the most popular methods for determining levels of bilingualism was developed by the Foreign Service Institute (FSI)⁵ of the U.S. Department of State. This method measures verbal and written performance to assess the functional bilingual ability of the individual being tested.

Since other characteristics (such as education, age, sex, education, etc.) often influence bilingualism, these characteristics must be considered in a study of bilingual ability. Blair (1990:53) indicates that the most important factors

⁵ Although the current acronym for the FSI is the ILR (for Interagency Language Round table), the more broadly used FSI acronym is used throughout this thesis.

influencing bilingualism are motivation and contact. In this context, contact means interaction with the second language which is closely related with age, education, occupation, sex, frequency of contact with mother-tongue speakers of the second language, etc. The age of the individual may relate to bilingual ability in at least three ways.

- Bilingualism may increase with age,
- Bilingualism may decrease with age, or
- Bilingualism may peak about middle age.

In the same way, gender may relate to bilingualism in at least three ways depending on one's social roles in the community.

- Women may be more bilingual than man,
- Men may be more bilingual than women, or
- Both may have about the same bilingual level.

Education is more complicated than age or gender because there are other minor factors involved. Blair (1990:56-57) presents several possibilities of bilingualism that relate to education, which depend on whether the second or third language is a language used in education. Since a third language is not in focus in this thesis, it will not be mentioned here. When a second language is a language of education, the level of bilingualism will increase with education. If an individual has little or no contact with a second language speaker, bilingual ability begins at a very low point while those who have more contact with second language speakers will have more bilingual ability regardless of the level of education.

Blair (1990:97-106) also introduced self-evaluation questionnaires to determine an individual's ability based on how much she can accomplish using the second language. This questionnaire is sometimes called a 'Can do questionnaire.' This type of questionnaire presents different situations requiring different levels of language ability to successfully accomplish them. A subject will report whether she is able to accomplish the task or not. This questionnaire employs 'Yes/No' questions. There are different levels in the questionnaire indicating different

degrees of proficiency in the second language. Each level contains a few questions. In order to score the questionnaire, the subject has to give a positive answer to every question at that level. If they don't give all positive responses to that level, they are assigned in the next lower level to which they gave all positive responses. The questions used in this survey are shown in the next chapter.

PAYAP UNIVERSITY