CHAPTER 3

METHODOLOGY OF SAMPLING

3.1 INTRODUCTION

The sociolinguistic dialectology approach involves major shifts in basic assumptions and procedures of traditional dialectology, population sampling, linguistic sampling, technique of data collection and presentation of the data (Francis, 1983). This chapter discusses the methodology of sampling, which includes the selection of the focal area, informants and variations in language to be investigated**.**

3.2 FOCAL AREA

Kota Kinabalu was chosen as the research area or focal area for this study. It is the capital of Sabah, East Malaysia, on the island of Borneo. It is located at latitude 5.98333° and longitude 116.06667°.

Kota Kinabalu, or Jesselton as it was formerly known, began its history before its official founding in 1899. It was also once known by various names: Deasoka, Singgah Mata and Api-Api. Deasoka is a Bajau word meaning "beneath a coconut tree" (Sabah State Government, 2000:31). In 1899 when the first works began on the building of Jesselton, there was an area called Deasoka to the south of the present city of Kota Kinabalu, which had a village and a coconut plantation. In Malay, the name, Singgah Mata means "a place upon which the eye lingers" *(ibid).* Apparently, Jesselton was an attractive place even then. Api-api is a much older name than Jesselton. It was mentioned in 1882 in the *British North Borneo Herald,* in reference to

a small village near the headland of Tanjung Aru. The name derived from the *Api-api* tree or *Avicennia*, which grew in abundance along the coast during that period. While Deasoka and Singgah Mata are no longer used, Api-Api or Api is still being used, especially by the Chinese.

The new township in Gaya Bay was named Jesselton in 1899 after Sir Charles Jessel, the Vice-Chairman of the British Chartered Company. The name Jesselton was however rarely used except for official purposes or in official documents. In 1967, Jesselton was renamed Kota Kinabalu, after the highest mountain in Southeast Asia, Mount Kinabalu. Today the name is commonly abbreviated to 'K.K.'. The name 'Kinabalu' is believed to originate from the Kadazandusun expression 'Aki Nabalu' where 'Aki' means grandfather and 'Nabalu' refers to eternity or the eight spiritual realms above the earth in the native belief system (*ibid*). Other theories suggest that Kinabalu means either "Chinese Widow" or "revered place of the dead" (http://en.wikipedia.org/wiki/Kota_Kinabalu).

Kota Kinabalu has come a long way since its humble beginnings as a trading and administrative centre of the British Chartered Company in the late 19th century. Since then, it has become a thriving seaport, an economic centre, a political centre and a modern gateway to the rest of Sabah. Today, Kota Kinabalu is the sixth town in Malaysia to achieve the historical milestone of becoming a city on February 2, 2000.

The city of Kota Kinabalu encompasses 9 zones covering 350,702 square km. The zones are Luyang, Sembulan, Tanjung Aru, Kepayan, Likas, Inanam, Menggatal, Telipok and the Central Business District as shown in Figure 3.1.

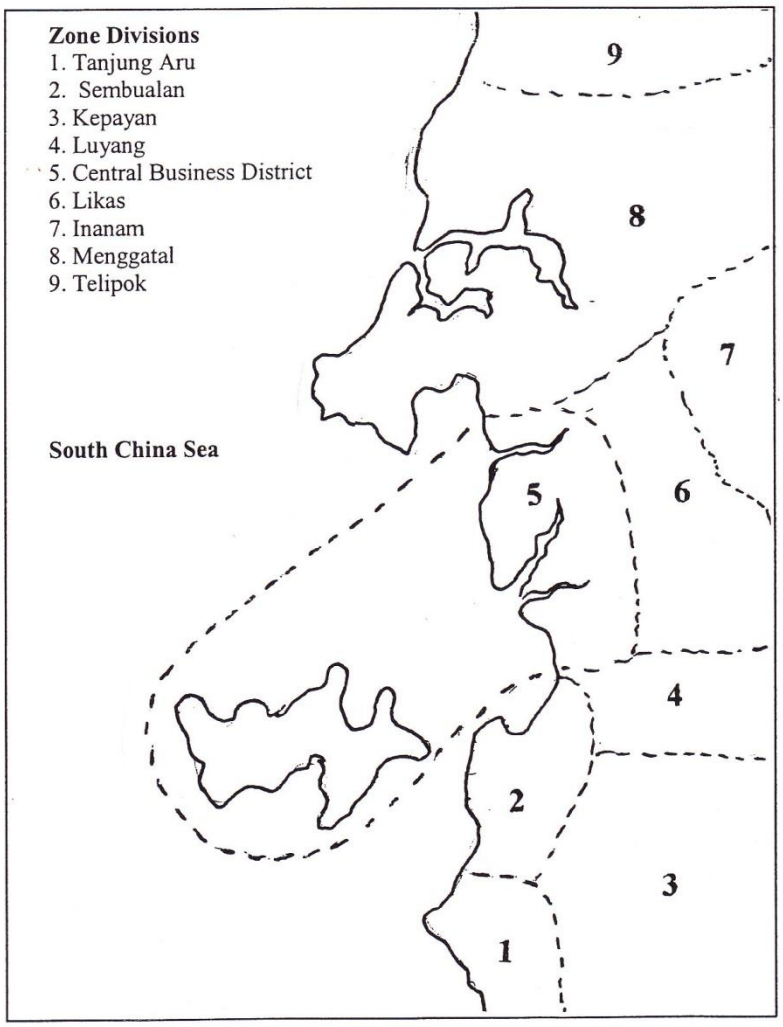

Figure 3.1 Zone Divisions of the City of Kota Kinabalu

People from all over Sabah come to Kota Kinabalu for business, education and employment. The population of Kota Kinabalu was 372.047 in 2000 (*Yearbook of Statistics Sabah 2005*) with an average annual growth rate of 2.42% making it 452,940 in 2010 (*Yearbook of Statistics Sabah 2010*). There are many different ethnic groups and immigrants living in Kota Kinabalu as shown in Table 3.1.

Table 3.1: The Population of the City of Kota Kinabalu

Ethnicity	**Population**	**Percentage**
Malay	49,609	13.33%
Kadazandusun	60,715	16.32%
Bajau	60,857	16.36%
Murut	2,997	0.81%
Other Bumiputeras	43,243	11.62%
Chinese	85,297	22.93%

Table 3.1, cont.

Ethnicity	Population	Percentage
Other Non-Bumiputeras	11,762	3.16%
Non-Citizen	57,567	15.47%
Total	**372,047**	**100.00%**

Source: Population Distribution and Basic Demographic Characteristics-Population and House Census of Malaysia, 2000 in Yearbook of Statistics Sabah 2010

From Table 3.1, it can be concluded that the Chinese are the largest ethnic group living in Kota Kinabalu, followed by Bajau, Kadazandusun and Malay. Hence, for the purpose of this study, Kota Kinabalu is a strategic focal area for linguistic research given its numerous Malay and non-Malay ethnic groups. Furthermore, the Sabah Malay dialect (SMD) is an urban variety, which emerges in urban areas with a good mixture of Malay and non-Malay elements. With this in view, Kota Kinabalu can be considered as a most suitable area for this type of study.

3.3 THE INFORMANTS

Informants are the most important element in a dialectology study. Hence, informant sampling is crucial where the size of sampling and method of sampling have to be decided on. Moreover, the appropriate ethnography criteria such as socio-cultural background and socio-economic background of the informants have to be determined for this study.

3.3.1 SAMPLING OF INFORMANTS

The key word for informant sampling in sociolinguistic dialectology is ‘representative’, in which informants selected represent the population. Hence, great care was employed by the researcher in choosing informants. Preparation and planning were devoted to the identification and selection of a representative cross-section of the population in the focal area. A proportionate distribution of social factors or

parameters such as social stratification, age, gender and ethnic membership was also carefully determined in the study.

Informant sampling involves two major decisions. First, the number of informants, which is the size of the sample, and second, the procedure of selecting the informants, which is the method of sampling. For this study, the researcher examined informant sampling methods of Labov on the Lower East Side of New York City (1966), Trudgill on the city of Norwich (1974), the local pioneering work of Wong (1987) on Kampung Kerinchi and Idris bin Aman (1995) on Shah Alam. Each one of them has been adapted where applicable.

3.3.1.1 SIZE OF SAMPLING

Kota Kinabalu covers an area of 350,702 square km with a population of 452,940 (*Yearbook of Statistics Sabah 2010*). By adapting Labov's study in the Lower East Side, where he eliminated "those who had moved away, and those who lived less than two years in the area" (1966: 157), the researcher chose only those who had been residing in Kota Kinabalu for more than two years to be the informants.

The numbers of informants used in previous urban dialectology studies, were ranged from 0.03- 0.14 % of the total population in the selected area. Labov's sample comprised of 195 informants which was about 1:690 or 0.14% of the population of Lower East Side; Trudgill's consisted of 50 which was about 1:3,200 or 0.03% of the population of Norwich, while Wong's was 72, about 1:42 or 0.14% of the Kerinchi's population. Idris, on the other hand, had 32 informants which was about 1:4,062 or 0.03% of Shah Alam's population. For this study, 120 informants were involved, i.e

1:2,961 or 0.04% of the Kota Kinabalu's population[1] (*Yearbook of Statistics Sabah 2005*) at the time the fieldwork was conducted. This percentage is slightly above the minimum percentage of 0.03%, used by previous dialectology studies.

3.3.1.2 METHODS OF SAMPLING

For this study, the researcher adapted the sampling methods used by Wong (1987) and Idris bin Aman (1995). A quasi-random method was used in order to select informants based on house numbers and individual name lists.

The national election list was not applied because people might have shifted or moved from their last recorded addresses. An impromptu method was not applicable because the questionnaire and interview would be lengthy and it may have taken 20 to 40 minutes for each informant to respond. Using a method such as the one based upon house numbers by itself would not be feasible for the city of Kota Kinabalu. This is because the housing development in Kota Kinabalu is not as well planned as it is in Shah Alam, where Idris conducted his research. Furthermore, most of the informants available should this method have been to be applied, would be limited to housewives, senior citizens and school children, as they are the ones who are most likely to be at home. On the other hand, the name list method is feasible since it could be obtained from office payrolls, school registers, university registrars and others. It was then decided to use a combination of the sampling methods, one that is based on house number and the other based on individual name lists.

[1] This number of informants is also equivalent to 1:3,774 or 0.037% of the current Kota Kinabalu's population (*Yearbook of Statistics Sabah 2010*), which is also above the minimum percentage. As the fieldwork and interviews are conducted in 2006, thus the Statistics Sabah 2005 was referred as the selection of informants was made.

For informant sampling based on house number, a few zones in Kota Kinabalu were selected; these are Inanam, Menggatal and Telipok to represent the focal area as these zones are more organised in housing set-up. To quote Moser (1958:76):

> Let us suppose that a sample of individuals is to be selected in a town with a population of 250,000. Most probably, one would decide to concentrate the interviews in a few areas, so the first step may be to pick some of the wards in town. (As cited in Trudgill, 1974a:22)

Following this, a few housing areas from each zone were selected. Next, a few houses were selected from each of these housing areas.

Using the quasi-random method, lists of selected housing areas were compiled beforehand. This method is used when the population of the focal area is too large. Houses were selected based on numbering. If an area had 200 houses and the researcher needed 10 informants, the mean would be 20. Houses were then selected at every count of 20 based on the mean. For example, when the researcher chose house No. 5 to be the first house, the following houses to be selected would be 25, 45, 65, 85, 105, 125, 145, 165 and 185. House members that fulfilled the age requirement and residential duration qualified to become informants of the study.

The second method of informant selection was based on a selection of individual name lists from several places. Places or venues of the interview were again determined earlier by adopting the quasi-random method, where certain government departments, private offices, schools, universities and supermarkets were selected. These included Universiti Malaysia Sabah, SMK Tebobon, SMK Bandaraya, Yayasan Sabah, Likas Mart, Menggatal Market, Inanam Market, Parkson Grand, Multibake and the Filipino Market. Individual name lists of employees and students were collected

earlier and selected in advance before the interviews. For example, Universiti Malaysia Sabah became one of the venues for the study. Name lists of students who reside in Kota Kinabalu were obtained by sorting them according to their permanent address. If there were 200 students from Kota Kinabalu and the researcher needed 20 informants, the mean would be 10. Similar to the house number method, students were selected at every count of 10 based on the mean, for example, if the first student is no 3 the next student would be no 13 and so forth. The same method was applied to the selection of informants from other pre-determined venues.

3.3.2 ETHNOGRAPHY OF INFORMANTS

For representative sampling, informants need to be chosen from different socio-cultural background and socio-economic backgrounds. This representative group of informants chooses and uses language differently from one another in their daily lives.

3.3.2.1 SOCIO-CULTURAL BACKGROUND

In this study, the socio-cultural background of informants include their gender, age and ethnic background, residence, hometown or country of origin. Informants must be residing within the zones of Kota Kinabalu. Table 3.2 shows the distribution of their residence in Kota Kinabalu.

Table 3.2: Area of Residence of Informants

Area of Residence	**No. of Informants**	**% of Informants**
Luyang	4	3.33
Sembulan	1	0.83
Tanjung Aru	1	0.83
Kepayan	13	10.83
Likas	13	10.83
Inanam	7	5.83
Menggatal	56	46.67
Telipok	11	9.17
KK Central Business District	14	11.67
Total	**120**	**100.00**

Kota Kinabalu is a city of opportunities. Most informants come from other districts and some even from other countries for career advancement, education and business opportunities. For example, out of 120 informants, only 19 informants were originally from Kota Kinabalu, 84 informants from other parts of Sabah, 2 from Sarawak, 3 from Peninsular Malaysia, 4 from Indonesia, 4 from the Philippines, and 4 from Japan as illustrated in Table 3.3.

Table 3.3: Hometown/Home Country of Informants

Hometown/ Home country	**No. of Informants**	**% of Informants**
SABAH	103	90.00
Beaufort	9	7.50
Keningau	3	2.50
Kota Belud	5	4.17
Kota Kinabalu	19	15.83
Kota Marudu	4	3.33
Kuala Penyu	2	1.67
Kudat	2	1.67
Kunak	1	0.83
Labuan FT	1	0.83
Membakut	5	4.17
Papar	7	5.83
Penampang	5	4.17
Pitas	1	0.83
Putatan	4	3.33
Ranau	10	8.33
Sandakan	5	1.67
Semporna	2	1.67
Sipitang	1	0.83
Tambunan	2	1.67
Tamparuli	1	0.83
Tandek	1	0.83
Tawau	7	5.83
Telipok	1	1.67
Tenghilan	2	1.67
Tenom	1	0.83
Tuaran	2	1.67
SARAWAK	2	1.67
PENINSULAR MALAYSIA	3	2.50
INDONESIA	4	3.33
THE PHILIPPINES	4	3.33
JAPAN	4	3.33
Total	**120**	**100.00**

According to the 2005 Statistics, 50.65% of Kota Kinabalu population is male and 49.36% female. Hence, for this study, an equal or almost equal number of males and females was the target. Out of 120 informants, 52 of them are male and 68 of them are female. The percentage of both genders is shown in Table 3.4.

Table 3.4: Gender of Informants

Gender	No. of Informants	% of Informants	% Statistics 2005
Male	52	43.33	50.65
Female	68	56.67	49.36
Total	**120**	**100.00**	**100.00**

The informants of the study were between the ages of 15 and 64 years old. Based on the *Yearbook Statistics of Sabah (2005:19),* a five-year age grouping was used; the youngest age group of this study was 15-19 year olds, whilst the oldest was between 60-64 year olds. Almost every age group had a higher percentage of representatives in contrast with the age grouping in Statistics 2005 with the exception of the youngest age group of 15-19 year olds, which only represents 10.00% of total informants, but 11.50% of the total population of Kota Kinabalu. The percentage of the age group variable of this study as contrasted to the Statistics 2005 list is shown in Table 3.5.

Table 3.5: Age of Informants

Age Group	No. of Informants	% of Informants	% Statistics 2005
15-19 year olds	11	9.17	11.50
20-24 year olds	27	22.50	10.33
25-29 year olds	19	15.83	8.52
30-34 year olds	10	8.33	7.87
35-39 year olds	24	20.00	7.97
40-44 year olds	8	6.67	6.54
45-49 year olds	7	5.83	4.61
50-54 year olds	3	2.50	2.83
55-59 year olds	3	2.50	1.80
60-64 year olds	8	6.67	1.35
Total	**120**	**100.00**	**63.32**

The informants are from various ethnic groups. The largest ethnic group represented is Kadazandusun which comprises 39 informants, followed by Bajau 24, Malays 18, Bugis 11, Chinese 9 and other ethnic groups a total of 19. However, 22 informants (18.33%) are of mixed ethnic parentage. The classification and distribution of the other ethnic groups is illustrated in Table 3.6.

Table 3.6: Ethnic Membership of Informants

Ethnic Group	**No. of Informants**	**% of Informants**
Malay	4	3.33
Malay + Kadazan	1	0.83
Brunei	9	7.50
Brunei + Kedayan	1	0.83
Kedayan	3	2.50
Kadazan	7	5.83
Kadazan + Dusun	1	0.83
Kadazan + Bidayuh	1	0.83
Dusun	17	16.67
Dusun + Kadazan	6	2.50
Dusun + Brunei	1	0.83
Dusun + Bugis	1	0.83
Dusun + Timor	1	0.83
Sungei	1	0.83
Tombonuo	1	0.83
Bisaya	2	1.67
Bajau	19	15.83
Bajau + Dusun	3	2.50
Bajau + Malay	1	0.83
Bajau + Timor	1	0.83
Murut	1	0.83
Bugis	9	7.50
Bugis + Bajau	1	0.83
Bugis + Jawa	1	0.83
Kelabit	1	0.83
Tidong	2	1.67
Banjar	1	0.83
Iranun	1	0.83
Chinese	7	5.83
Sino-Kadazan	2	1.67
Indian	1	0.83
Indonesian	4	3.33
Filipinos	4	3.33
Japanese	4	3.33
Total	**120**	**100.00**

3.3.2.2 SOCIO-ECONOMIC BACKGROUND

The socio-economic background of the informants is determined by occupation, income, highest level of education and type of housing they dwell in. These determinants are the components of the informant's Social Stratification (3.4.1.4).

In terms of occupation, the informants' jobs range from cleaners to government officers. Some informants are pensioners and housewives, whereas the others are students. The informants' occupations are shown in Table 3.7.

Table 3.7: Occupation of Informants

Occupation	No. of Informants	% of Informants
Technician	2	1.67
Contractor	1	0.83
Officer	6	5.00
Assist. Officer	4	3.33
Senior Admin. Assist.	3	2.50
Admin. Assist.	8	6.67
General Assist.	2	1.67
Laboratory Assist.	1	0.83
Cleaner	2	1.67
Maid	1	0.83
Researcher	2	1.67
Operator	1	0.83
Driver	4	3.33
Senior Lecturer	2	1.67
Lecturer	11	9.17
Language Teacher	6	5.00
Teacher	4	3.33
Tutor	1	0.83
Gardener	4	3.33
Farmer	2	1.67
Guard	1	0.83
Salesperson	2	1.67
Hawker	6	5.00
Waiter/waitress	2	1.67
Tailor	1	0.83
Supervisor	1	0.83
Own Business	3	2.50
Housewife	4	3.33
Student	30	25.00
Pensioner	3	2.50
Total	**120**	**100.00**

Where income is concerned, categorisation is based on the monthly income of the informants by a one thousand Malaysian Ringgit count starting from RM500 a month as illustrated in Table 3.8. About 28 informants, who are students, do not have a substantial income. They only have a student allowance.

Table 3.8: Income of Informants

Monthly Income (RM)	**No. of Informants**	**% of Informants**
RM 5,500 and above	1	0.83
RM 4,500 - 5,499	5	4.17
RM 3,500 - 4,499	7	5.83
RM 2,500 - 3,499	13	10.83
RM 1,500 - 2,499	20	16.67
RM 500 - 1,499	24	20.00
RM 499 and below	16	13.33
Student Allowance	34	28.33
Total	**120**	**100.00**

In terms of education, there are 31 informants who are university graduates, 18 college graduates, 50 secondary school leavers, 9 primary school leavers and 12 have no formal schooling. The percentage is shown in Table 3.9.

Table 3.9: Education Level of Informants

Education Level	**No. of Informants**	**% of Informants**
University	31	25.83
College	18	15.00
Secondary School	50	41.67
Primary School	9	7.50
No Formal Schooling	12	10.00
Total	**120**	**100.00**

The informants live in various types of houses ranging from squatter huts to bungalows. Most informants live in terrace houses, which are double- or single-storeyed. Informants' types of houses are shown in Table 3.10.

Table 3.10: Housing of Informants

Type of Houses	No. of Informants	% of Informants
Bungalow/Detached	6	5.00
Double/Single-Storey Terrace	42	35.00
Apartment/Flat	18	15.00
Kampung House	49	40.83
Squatter Huts	5	4.17
Total	**120**	**100.00**

3.3.2.3 LINGUISTIC BACKGROUND

Although SMD could have begun as a common language of interaction used by non-Malays with the Malays or vice-versa in market places, today its function has extended to several other settings owing to many factors. Among these factors are diversity of the community and the great number of inter-ethnic marriages. Due to mixed-parentage, most couples of inter-ethnic marriages have opted to use SMD as home language[2] when speaking with their spouse at home and later to their children. Consequently, SMD has become the first language[3] for a large number of mix-parentage children. Some children grow up with multilingual competence; they speak one or two of their mother tongues[4] alongside SMD. SMD is not used widely as a tool of communication between different ethnic groups in market places, but is also used in government or private offices, and in formal and non-formal situations. Today, there is

[2] Home language is the language that is used at home to speak with spouse and children, especially in inter-ethnic marriage setting. It may be the language of the mother or the father or neither. However, whatever is the home language for the parents may most probably end up to be the L1 for the children. The choice could be dominance, environmental or pragmatics. For example, one parent is a Dusun and speaks Dusun as L1, another parent is a Chinese who speaks fluent Chinese. They may decide either Dusun or Chinese to be the language of interaction between them, depending on which side of the family is more dominant in their extended family life. It also depends on the area that the couple live in, e. g Ranau which is predominantly a Dusun area and Sandakan which is predominated by Chinese. But in most cases, a couple belonging to different ethnic groups will opt for SMD to be their home language. Hence the children will grow up speaking only SMD alongside Dusun or/and Chinese as their L1.

[3] For this study, first language or L1 is refers to the language that a person acquired first in childhood and is spoken in interaction with his/her parents or siblings at home. It may or may not be the person's mother tongue or native language.

[4] Mother tongue also refers to an ancestral language. It is the L1 of a person's parents, but not necessary his/her L1. For example, Dusun, can be the mother tongue of someone, but due to certain factors, parents choose to speak SMD to their children at home. Hence, Dusun is the children's mother tongue or ancestral language but not their L1. In this case, SMD is the L1.

an increasing number of local radio stations using fully SMD as a medium of communication, such as Sabah Suria FM and KKFM.

Based on the data collected in Questionnaire Part I-Section B on language choices and use (see Appendix Ai), it can be summarised that SMD plays an important role in the daily lives of the informants as shown in Table 3.11. Today, the initial function of SMD as a lingua franca can be seen in the language choices and use of the informants. Of 120 informants, 104 (86.66%) use SMD to interact with other Sabahans of different ethnicities, while only 9 informants (7.50%) use Standard Malay. About 82 informants (68.33%) use SMD to interact with Sarawakians or Peninsular Malaysians, while only 23 informants (19.16%) use Standard Malay. About 99 informants (82.50%) use SMD when interacting with other non-Malaysians such Indonesians, Filipinos and other immigrants in Sabah. In contrast only 8 informants (6.67%) use Standard Malay and 4 informants (3.33%) use English. However, 62 informants (51.66%) use SMD, and 29 informants (24.17%) use SMD along with other languages when interacting with people of their own ethnicity. Only 20 informants (16.70%) use their mother tongue or ancestral language when interacting with their own people.

Table 3.11: Language Choices and Use among Informants-Lingua Franca

Language Choices Language Functions		**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD and ENG**	**O/LG**
Lingua Franca with									
• Same Ethnic Group Sabahans (120)	#	62	2	20	0	2	25	2	1
	%	51.66	1.67	16.70	0.00	1.67	20.83	1.67	0.83
• Other Ethnic Group Sabahans (120)		104	9	0	0	0	0	7	0
		86.66	7.50	0.00	0.00	0.00	0.00	5.83	0.00
• Other Malaysians (120)		82	23	0	1	5	1	6	2
		68.33	19.16	0.00	0.83	4.17	0.83	5.00	1.67
• Other Non-Malaysians (120)		99	8	0	4	3	2	4	0
		82.50	6.67	0.00	3.33	2.50	1.67	3.33	0.00

One of the most common functions SMD is as lingua franca. It is expected to have limited functions as the language of trade, which Za'ba refers to as the bazaar

variety (Za'ba 1965), used to communicate in the market places and with vendors (Asmah Haji Omar, 1980). SMD is a lingua franca in local trading within Sabah (Wong, 2000). It is the lingua franca in the market places between vendors and buyers in shopping complexes between sales persons and customers; in restaurants between waitresses and patrons; and in public transport vehicles between drivers and passengers. Of the 120 informants, 96-108 informants (80-90%) speak SMD to vendors, salesmen, waiters, waitresses, bus drivers and conductors at one time or another.

Table 3.12: Language Choices and Use among Informants-Bazaar Language

Language Choices **Language Functions**	**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD and ENG**	**O/LG**
Speak with								
• Vendors (120) #	100	3	3	0	2	11	1	0
%	83.33	2.50	2.50	0.00	1.67	9.17	0.83	0.00
• Salesmen (120)	99	5	3	0	5	5	1	2
	82.50	4.17	2.50	0.00	4.17	4.17	0.83	1.67
• Waiters (120)	96	10	3	0	5	4	1	1
	80.00	8.33	2.50	0.00	4.17	3.33	0.83	0.83
• Drivers (120)	108	8	0	0	2	2	0	0
	90.00	6.67	0.00	0.00	1.67	1.67	0.00	0.00

According to the *Yearbook of Statistics Sabah 2005*, there are about 294,833 Malays in Sabah (including Labuan Federal Territory) who speak some varieties of Malay as their mother tongue. They are the native speakers of varieties of Malay subgroups, namely Cocos, Brunei and Kedayan, which are distinct from SMD. Although SMD is not considered a 'native language' of the Malay subgroups in Sabah, it is spoken as the first language by many Sabahans. One parent's linguistic background is a major determinant of language choices. Children of mixed parentage often speak SMD as their first language (Wong, 2000). Of the 120 informants in this study, 37 informants (30.83%) speak SMD to both their parents, 13 informants speak SMD alongside their mother tongue to their father (10.83%), and 12 to their mother (10%). The number of informants speaking SMD to siblings increases to 43 informants

(35.83%). Another 13 informants (10.92%) use SMD alongside their mother tongue and 4 informants (3.36%) use SMD alongside English when speaking with their siblings.

Table 3.13: Language Choices and Use among Informants-First Language

Language Choices Language Functions		**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD and ENG**	**SMD and O/LG**	**O/LG**
First Language with										
• Father (120)	#	37	4	61	1	1	13	2	1	0
	%	30.83	3.33	50.83	0.83	0.83	10.83	1.67	0.83	0.00
• Mother (120)		37	4	64	1	0	12	2	0	0
		30.83	3.33	53.33	0.83	0.00	10.00	1.67	0.00	0.00
• Siblings (119)		43	5	51	2	0	13	4	0	1
		36.13	4.20	42.85	1.60	0.00	10.92	3.36	0.00	0.84

As inter-ethnic marriage is a common practice in Sabah, where 18.33% (see 3.3.2.1) of the informants are of mixed parentage, they tend to speak SMD instead of their mother tongue to their spouses and later to their children. Hence, SMD is the home language to them. Of the 75 married informants, 35 (46.66%) speak SMD and 12 (10%) speak SMD alongside one or more languages to their spouse. Out of 68 informants who have children, 35 informants (51.47%) speak SMD and another 14 informants (11.66%) speak SMD alongside one or more languages to their children. This scenario is also noted by Saidatul Nornis Haji Mahali (2003:301):

> *Perkahwinan campur telah memudahkan laluan integrasi kerana setiap pasangan yang kahwin campur berusaha untuk mempelajari bahasa pasangan masing-masing di samping menggunakan bahasa Melayu (SMD) sebagai bahasa pengantar. Hasilnya, anak-anak yang dilahirkan akan menguasa lebih dari satu bahasa sukuan dan fasih menggunakan bahasa Melayu (SMD) sebagai alat perhubungan sosial*
> (Intermarriage has facilitated the integration path for each pair of mixed marriage couples trying to learn their language while using the Malay language (SMD) as a medium. As a result, children born will capture more than one ethnic and fluent language Malay language (SMD) as a means of social communication).

Consequently, inter-ethnic marriage decreases the use of mother tongue as a home language. From 64 informants (53.33%) who speak their mother tongue as their first language, only 20 informants (26.66%) speak their mother tongue to their spouses and only 8 informants (11.76%) to their children. As a result, only the children of these 8 informants use their mother tongues as their first language, whereas about half of the informants (51.47%) use SMD as the language of interaction with their children. As these children grow up speak SMD as their first language, they will continue this tradition in bringing up their children later.

Table 3.14: Language Choices and Use among the Informants-Home Language

Language Choices Language Functions		**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD and ENG**	**SMD and O/LG**	**O/LG**
Home Language with										
• Spouse (75)	#	35	4	20	3	1	6	4	1	1
	%	46.66	5.33	26.66	3.99	1.33	8.00	5.33	1.33	1.33
• Children (68)		35	5	8	3	1	6	3	4	3
		51.47	7.35	11.76	4.41	1.47	8.82	4.41	5.88	4.41

Even though STM is the official language of government administration, the medium of instruction in schools, and the language of mass media in Sabah since 1967, SMD is still widely used. Even in formal situations, SMD is frequently used. For instance, between officers and clients in government offices, about 62 informants (57.40%) use SMD and only 25 informants (23.15%) use STM. In school, between students and teachers or parents and teachers, 60 informants (52.63%) use SMD and only 24 informants (21.05%) use STM.

Table 3.15: Language Choices and Use among the Informants-Formal Language

Language Choices Language Functions		**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD And ENG**	**SMD and O/LG**	**O/LG**
Formal Language with										
• Officers (108)	#	62	25	0	2	5	1	3	1	9
	%	57.40	23.15	0.00	1.85	4.62	0.92	2.78	0.92	8.33
• Teachers (114)		60	24	0	4	4	1	8	4	10
		52.63	21.05	0.00	3.50	3.50	0.88	7.01	3.50	8.77

For language choices and use in government and private offices, SMD is even spoken extensively between officers. Among working informants, 83.33% speak SMD along with other languages to their colleagues and subordinates. The percentage of informants using SMD decreases to 60.67% when communicating with superiors. This indicates that SMD is spoken mostly in informal situations. Conversely, the more formal the situation, the more appropriate STM becomes. Only 7.29% uses STM when speaking to colleagues. The percentage increases to 9.52% when speaking to subordinates. It increases to a greater percentage of 22.47 when speaking to superiors.

Table 3.16: Language Choices and Use among Informants-Office Language

Language Choices **Language Functions**	**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD and ENG**	**SMD and O/LG**	**O/LG**
Office Language with									
• Colleagues (96) #	45	7	4	1	6	11	7	11	4
%	46.88	7.29	4.17	1.04	6.25	11.46	7.29	11.46	4.17
• Superiors (89)	41	20	0	5	4	1	6	2	10
	46.06	22.47	0.00	5.62	3.33	1.12	6.74	2.24	11.24
• Subordinates (84)	60	8	2	1	6	3	1	0	3
	71.43	9.52	2.38	1.19	7.14	3.57	1.19	0.00	3.57

Besides that, SMD is also a choice in interaction between friends. Out of 120 informants, 62 (51.66%) use SMD to communicate with their friends of the same ethnicity, while 104 informants (86.66%) use SMD to interact with friends of other ethnicities. 82 informants (68.33%) use SMD to communicate with other Malaysians, while 99 informants (82.50%) use SMD to communicate with non-Malaysians, including Indonesians and Filipinos. It is indeed a language of solidarity among various speakers.

Table 3.17: Language Choices and Use Among Informants-Language of Solidarity

Language Choices **Language Functions**	**SMD**	**STM**	**M/T**	**ENG**	**SMD and STM**	**SMD and M/T**	**SMD And ENG**	**SMD and O/LG**
Lg. of Solidarity with								
• Same Ethnic Group #	50	2	39	0	1	24	4	0
Friends (120) %	41.66	1.67	32.50	0.00	0.83	20.00	3.33	0.00
• Other Ethnic Group	96	11	0	2	3	0	6	1
Friends (120)	80.00	9.17	0.00	1.67	2.50	0.00	5.00	1.67

3.4 VARIATIONS IN LANGUAGE

This study observes and groups linguistic variations into three principal categories, namely linguistic variations, social variations and stylistic variations. These language variations are interrelated as there are bound to be linguistic differences, such as phonological differentiation in social context, where individual speakers use variations according to those used by the group they belong to, determined by age, gender, ethnic membership and socio-economic level, as well as in different speech style ranging from formal to informal speech.

3.4.1 SOCIAL VARIATIONS

The incidence of language variations in society could be influenced by many social parameters. Nonetheless this study discusses only four parameters, namely age, gender, ethnic membership and social stratification.

Each of the variables chosen for this study is classified into smaller groups or variants. However, the classification here is different from those of the informants' socio-economic background in 3.3.2.2. Here, each variant of all the social variables, namely age, gender and ethnicity, is given a numerical index code (see Appendix Aii). These numerical index codes are only regarded as indicators to identify the informants' social groupings and they are not assigned any significant value. For example, under the gender variable, the male variant is given a code of 1, while the female variant is given a code of 2. However, for the social stratification, each variant of all the variables is given a numerical index score, which carries a certain value. The numerical index scores from all four indices, i.e. Occupation, Income, Education and Housing is added to obtain the social stratification index code (see Table 3.17). The categorisation of each grouping for every variable is further discussed below

3.4.1.1 GENDER

In most societies, individuals tend to associate more with members of their own gender than with those of the opposite gender due to economic, social and religious reasons. This segregation leads to various degrees of linguistic variations. In most studies, one gender tends to use more of the higher status variants than the other. Therefore, this study investigates how males and females influence linguistic variations of SMD. Each gender is given a numerical index code; in which male is encoded as 1 and female is encoded as 2. This gender classification and the number of informants in each group are shown in Table 3:18.

Table 3.18: Gender Index

Gender Group	No. of Informant	Code
Female	68	2
Male	52	1

3.4.1.2 AGE

A society can be stratified by age. Individuals pass through several age groups before attaining social and linguistic maturity. Anyone between the age of 15 and 64 is eligible to be an informant of this study. For this study, the informants are divided into five different age groups: i) 15-24 year olds; ii) 25-34 year olds; iii) 35-44 year olds; iv) 45-54 year olds; and v) 55-64 year olds. A numerical index code ranging from 1 to 5 is given to each grouping. The youngest age group is encoded as 5 and the oldest as 1. This age classification and the number of informants in each group are shown in Table 3:19.

Table 3.19: Age Index

Age Group	No. of Informant	Code
15-24 year olds	38	5
25-34 year olds	29	4
35-44 year olds	32	3
45-54 year olds	10	2
55-64 year olds	11	1

3.4.1.3 ETHNIC MEMBERSHIP

In many communities, different ethnic groups speak different languages. However, what is more interesting in this study is where different ethnic groups speak the same language but differ qualitatively or quantitatively in their use of particular variables (Chamber and Trudgill, 1980:74). This is particularly true in the case of SMD, where all Sabahans speak SMD in their daily lives at one time or another.

Sabah is a multi-ethnic state, which has as many as 41 ethnic and sub ethnic groups (Mat Zin Mat Kib, 2005). However, for this study, the researcher has divided the society into seven ethnic groups, namely Malay (MLY), Kadazandusun (KZD), Bajau (BJU), Bugis (BGS), Other Bumiputera (BMP), Chinese (CHN) and Other Non-Bumiputera (ONB). Murut is not categorised as an ethnic group[5] for this study because there is only one Murut informant; this informant is included in the Other Bumiputera (BMP) group. On the other hand, Bugis is studied as an ethnic group, as there are 11 informants in the study.

The definition of ethnic membership and ethnic grouping used in this study differs slightly from those of the Department of Statistics Malaysia or Department of Registration Malaysia. It is based on the linguistic grouping adapted from King and King (1984:328), rather than political or ethnographic grouping. This is chosen because ethnic membership is one of the social variables that plays a significant role in determining linguistic variables. For this study, the ethnic membership is classified as in Table 3.20:

[5] The Department of Statistics Malaysia (2005) divides the Sabah population into eight groups: Malay, Kadazandusun, Bajau, Murut, Other Bumiputera, Chinese, Other Non-Bumiputera and Non-Malaysian.

Table 3.20: Ethnic Classification

Ethnic Membership	Ethnic Subgroups
Malay (MLY)	Malay, Cocos Malay, Brunei Malay,Kedayan
Kadazan (KDZ)	Kadazan, Dusun, Kuijau, Tatana, Tambanua, Lotud, Sungai, Rungus, Bisaya
Bajau (BJU)	Bajau
Bugis (BGS)	Rappang, Bone, Enrekang, Wajo, Pinr
Other Bumiputera (BMP)	Murut, Jawa, Butung, Illanun, Suluk, Lundayeh, Bonggi, Ida'an, Banjar Tidong
Chinese (CHN)	All Chinese sub-groups
Other Non-Bumiputera (ONB)	Indian, Indonesian, Philipino, Japanese

Similar to the treatment of other social variables, each ethnic group is given a numerical index code ranging from 1 to 7; the MLY is encoded as 7, KZD as 6, BJU as 5, BGS as 4, BMP as 3, CHN as 2 and ONB as 1. This ethnic classification and the number of informants in each group are shown in Table 3.21.

Table 3.21: Ethnicity Index

Ethnic Membership	No. of Informant	Code
Malay (MLY)	19	7
Kadazan (KDZ)	38	6
Bajau (BJU)	24	5
Bugis (BGS)	10	4
Other Bumiputera (BMP)	7	3
Chinese (CHN)	9	2
Other Non-Bumiputera (ONB)	13	1

Although the above definitions serve as a guideline, it is not always easy to classify a person's ethnicity, especially in Sabah. The situation becomes even more complex with a high percentage of inter-marriages among different ethnic groups. As inter-marriages are common, it is quite rare to meet Sabahans who are descendants of one single ethnic group. For this study alone, about 18.33% of the informants are of mixed parentage (see 3.3.2.1).

There are Sino-Kadazans, who are of Chinese and Kadazan parentage. There are others with Bajau and Chinese parentage, Bajau and Dusun parentage, Brunei and Dusun parentage, Kedayan and Melayu parentage, and Murut and Kadazan parentage, just to name a few (refer to table 3.6). It would be less complex for this study if both parents belong to the same ethnic group such as Dusun and Kadazan**,** Brunei and Kedayan, Murut and Tagal or Bajau and Sama. This is because they belong to the same linguistic family: Kadazandusun, Malay, Murut and Bajau respectively.

Conversely, it is more difficult to define one's ethnicity if both parents are of inter-ethnic marriage and of different linguistic groups, for example the Sino-Kadazan, and those of Malay and Bajau parentage, Murut and Dusun parentage, Malay and Kadazan parentage or Chinese and Bajau parentage. For these groups, the researcher had to investigate further to find out which culture they follow more closely, for instance, which festivals they typically celebrate, the type of food they eat and the language they speak. It is important to link these informants to their cultural roots or ethnicity, as it influences their linguistic variations. If a person is of a mixed-parentage of Chinese and Kadazan or a Sino-Kadazan, who speaks little or none Kadazan, but uses Chinese on daily basic, it can be assumed that the Kadazan language would not influence his or her linguistic variations in the use of SMD, as much as the Chinese language would. Therefore, he or she is considered as a Chinese by this study.

3.4.1.4 SOCIAL STRATIFICATION

Social stratification or social class is abstract and complicated. This is particularly true in Malaysia which society is not stratified by social classes. Thus, in order to label anyone to belong to any social strata, an index of classification is needed for the purpose of the study.

3.4.1.4.1 INDEX OF CLASSIFICATION

In order to measure one's status or social class, sociologists have created various indices. Some argue that social class is determined by occupation (King and Raynor 1981); wealth or income, occupation and education (Horton and Hunt 1981); type of house that one lives in (Alias Baba 1992); occupation, income, education, housing, locality and father's occupation (Trudgill, 1974a).

For this study to be more objective and reliable, four indices were chosen to determine the social stratification of informants in Kota Kinabalu, namely education, occupation, income, and housing. Each variable of the social variations was then classified into different groups or variants for which numerical scores were given. In order to establish a social stratification index for each informant, index scores of all the four social variables were added together. For each of these indices, a scale ranging from 1 to 7 was assigned (see 3.4.1.4.1.1 - 3.4.1.4.1.4). It is possible for the informants to have a cumulative score ranging from 4 to 28. Based on the accumulated score from all four indices in this study, informants are then subdivided into different social classes ranking from the highest class to the lowest class. Since there is no existence of any social labels for Malaysian society, the researcher adapted and utilised Trudgill's methods of subdividing informants into six social classes were: Upper Middle Class (UMC), Middle Middle Class (MMC), Lower Middle Class (LMC), Upper Working Class (UWC), Middle Working Class (MWC) and Lower Working Class (LWC) (see 3.4.1.4.2). Similarly labeling was also employed by Idris bin Aman (1995). This classification does not represent the informants' standing in the community they belong to. They are only a classification used for the purpose of this research. Alternately, labeling such as Class 1, Class 2, Class 3, Class 4, Class 5 and Class 6 can be used to represent different hierarchy ranking from the highest to the lowest.

3.4.1.4.1.1 OCCUPATION

Occupation is what one does for a living. It is the most common and easily applied index to be used to classify social stratification or social class in the United Kingdom and the United States of America. This can also be applied in the case of Malaysia.

A seven-point scale was used to classify the type of occupations based on *Employment Statistics in Sabah 2002-2004 (Year Book of Statistics Sabah 2005:187)* as shown in Table 3.21. Each of these groupings was given a numerical index score ranging from 1 to 7. As shown in Table 3.22, at the top of the occupation index were Legislators, Senior Officials and Managers and each carries a score of 7. At the bottom of the scale were the Elementary Occupation which carried a score of 1.

Table 3.22: Occupation Index

Occupation	**Score**
Legislators, Senior Officials and Managers	7
Professionals	6
Technicians and Associate Professionals	5
Clerical Workers	4
Skilled Agricultural and Fishery Workers	3
Craft, Trades and Plantation Workers, andMachine-Operators	2
Elementary Occupation	1

A homemaker is not categorised by the occupation index; hence if an informant is a housewife the occupation of her husband is taken as a rough identification of her social stratification. If the informant is a student or a school leaver without a job, the occupation of his or her father or mother, whoever is the breadwinner in each case, was considered. For a pensioner, his or her last job before retirement was considered. For a senior citizen who had never been employed before, the occupation of his or her caretaker is taken into account. This is done in order to compensate for the absence of the informant's occupation.

3.4.1.4.1.2 INCOME

A person's standard of living is very much dependent upon his or her occupation and income. Wealth and income will determine one's lifestyle, especially in terms of activities, housing, dressing, food and other areas.

Similar to the occupation index, this income index is divided into a seven-point scale as shown in Table 3.23. Each of these groupings carries a numerical index score ranging from 1 to 7. For example, the highest scale of the income index which is RM5,500 and above is given a score of 7, while the lowest scale of income index which is RM860 and below is given a score of 1.

Table 3.23: Income Intake

Monthly Income (RM)	**Score**
RM 5,500 and above	7
RM 4,500 - 5,499	6
RM 3,500 - 4,499	5
RM 2,500 - 3,499	4
RM 1,500 - 2,499	3
RM 889 - 1,499	2
RM 888 and below	1

Since there is no standard classification of income, the researcher has adapted (with some adjustment) the income classification used by Idris bin Aman (1995). The bottom scale of the income intake, RM888, is based on the report on poverty eradication in Sabah in 2004 (*Ninth Malaysia Plan, 2006-2010*: 329).

Similar to the occupation variable, if the informant is a housewife, the income of her husband is taken into account. If an informant is a student, the income of each of his or her parents, whoever is the breadwinner in this case is considered. For a

pensioner, his or her pension is considered. For a senior citizen who has never been employed before, the income of his or her caretaker is taken into account.

3.4.1.4.1.3 EDUCATION

The education index in the study is adapted from Idris bin Aman (1995). It is divided into a seven-point scale as illustrated in Table 3.24. Each education grouping is given a numerical index score ranging from 1 to 7. For example, the Post Graduate Degree, a Masters Degree and Doctorate of Philosophy are at the highest scale of the education index and each carries a score of 7. The lowest scale of the education index, which encompasses primary school and no formal schooling, each carries a score of 1.

Table 3.24: Education Index

Education Level	Score
Post Graduate Degree/Masters/PhD	7
Bachelor Degree/Advance Degree	6
Certificate/Diploma	5
Form Six/STPM/STP/HSC	4
Form Five/SPM/MCE/SPVM/SC	3
Form Three/PMR/SRP/LCE	2
Primary School/UPSR/No Formal Schooling	1

3.4.1.4.1.4 HOUSING

Adapting Idris bin Aman's (1995) classification of housing index with some modification, this study utilises a housing index divided into a seven-point scale based on the type of house and ownership in the manner shown in Table 3.25. Each of these groupings carries a numerical score ranging from 1 to 7. At the top of the housing index are bungalows and detached houses owned by informants, each carrying a score of 7. At the bottom of the housing index are flat (rented), single–storey Kampong house (rented) and squatter (owned/rented), each carrying a score of 1.

Apart from detached and semi-detached houses, double and single storey terrace houses and flats, the researcher has added kampong houses and squatter houses into the Housing index. This is because not all houses are situated in well-organised and developed housing areas. There are still many kampong houses and even unnumbered houses such as squatter homes in Kota Kinabalu. Since bungalows and apartments are common, they are added to the index list as well.

Table 3.25: Housing Index

Type of House	Score
Bungalow/Detached House (owned)	7
Bungalow/Detached (rented)	6
Semi-Detached/Condominium (owned)	5
Semi-Detached/Condominium (rented)/ Double-storey terrace (owned)/	4
Double-storey terrace (rented)/ Single-storey terrace/Apartment / Double–storey Kampong house (owned)	3
Single-storey terrace/Apartment/ Double–storey Kampong house (rented)/Flat/Single–storey Kampong house (owned)	2
Flat/ Single–storey Kampong house (rented)/ Squatter	1

As for ownership, houses are divided into bought and rented categories. The definition of 'owned' refers to the houses that are owned or inherited by informants or any of their family members. The term "rented" refers to houses that are rented or provided by employers such as government quarters or employee quarters of private organisations.

3.4.1.4.2 CLASSIFICATION OF SOCIAL STRATIFICATION

All numerical index scores of the informant from all the four social class indices of education, occupation, income, and housing, are added up to determine the informant's social stratification index score and therefore the social class of the informant. For example, if an informant accumulates 20 scores, with scores of 5, 4, 6, 5 from the four indices respectively (3.4.1.4.1.1-3.4.1.4.1.4), he or she is classified as a Middle Middle

Class informant. The composition of social classes and the number of informants in each class can be seen in Table 3.26.

Table 3.26: Composition of Social Stratification Index

Social Stratification	Total Score	No. of Informant	Code
Upper Middle Class (UMC)	24-28	0	6
Middle Middle Class (MMC)	20-23	15	5
Lower Middle Class (LMC)	16-19	26	4
Upper Working Class (UWC)	12-15	25	3
Middle Working Class (MWC)	08-11	29	2
Lower Worker Class (LWC)	04-07	25	1

3.4.2 STYLISTIC VARIATIONS

Stylistic variations comprise speech contexts[6]. There are four well-known speech styles, which represent four clearly differentiated registers. These can be elicited and recognized by the identification of situations and contexts such as Casual Speech, Careful or Formal Speech, Reading Connected Text and Word-List (Labov, 1966, Trudgill, 1974a).

After determining each variable of the social variations as mentioned in 3.3.1, the researcher set each variable or speech context of the stylistic variations. Data collection was organized so that formality of speech could be measured ranging from most formal to least formal and included four variables as adapted from Labov (1966) and Trudgill (1974a), namely, Word-List Style (WLS), Reading-Passage Style (RPS), Careful or Formal Style (FS) and Casual Style (CS).

[6]The term stylistic variation used in this study differs from those registers used by Asmah Haji Omar in *Bahasa DiRaja* (1984, 2004), *Bahasa Surat Rasmi* (1983b), *Bahasa Iklan Perniagaan* (1984), *Bahasa Borang* (1985), *Bahasa Malaysia Saintifik* (1987) and *Bahasa Laporan* (1997) (see 2.3.2.2).

3.4.2.1 WORD-LIST STYLE (WLS)

Word-List Style (WLS) is the most formal form of speech on the scale (Trudgill 1974a). For this study, a compilation of 150 word-list from Swadesh word-list (Nothofer, 1991) and word-list for the use of Borneo was used (Prentice, 1969). The ideal approach to go about the word-list would be through an "indirect approach" where questions would be asked in such a way that the answers would provide the words expected by the researcher (McDavid 1974, Kurath 1972 and Shuy 1967). However, this approach needs 9-10 hours or 2-3 sessions for each informant. Due to time constraint the researcher opted for the 'direct approach' where informants were asked to read aloud the word-list word by word.

3.4.2.2 READING-PASSAGE STYLE (RPS)

The Reading-Passage Style (RPS) is less formal than WLS. For this style, informants were required to read two passages, in this case newspaper articles from a local newspaper *Harian Express*, which were seen to contain all possible phonological variations. The passages chosen were related to local issues such as the Tsunami alert in Sabah and the storm in Kota Kinabalu, which happened not long before the research was conducted. This is to ensure that the informants could read with a certain degree of comprehension.

3.4.2.3 FORMAL STYLE (FS)

Careful or Formal Style (FS) is less formal than WLS and RPS and it is the third in the formal speech scale. It is not difficult to obtain this kind of speech, especially when the informant is interviewed by a stranger and the conversation is being recorded. Here, the researcher asked more formal questions regarding the city of Kota Kinabalu,

the informant's house and their school or place of work. Informants were expected to give their opinions and suggestions on the topics discussed.

3.4.2.4 CASUAL STYLE (CS)

Casual Style (CS) is the least formal form of speech on the scale. It is difficult to obtain this kind of speech, as informants tend to be formal when they are interviewed. Labov (1966) suggests that casual speech can be obtained through spontaneous speech, as "spontaneous speech is the counterpart of casual speech which does occur in formal contexts, not in response to the formal situation, but in spite of it". He also suggests three methods of getting casual speech outside the interview context. The first method is before, after or during breaks in an interview. The second method is when informants are talking to a third party. And the last method is when the informants are talking about tragedies. Here, in this study, the researcher used the first and third methods. For example the casual speech questions were allocated in the last part of the questionnaire, when informants were assumed to be at their most relaxed state. The main questions asked were about tragedies and near death experiences of the informants such as the Tsunami alert 2004, which was frightening for most Sabahans. Questions about TV programmes such as reality shows, for example *Akademi Fantasia*, were also asked, as these shows are viewer-centred and are thought to provoke emotion.

3.4.3 LINGUISTIC PHONOLOGICAL VARIATIONS

The incidence of variation in language is to be found in all aspects of the language: lexical item, phonology, grammar and semantics. All aspects of language are subject to change in any language due to social factors, with resulting variations. However, this study focuses only on phonological differentiation of certain vowels and consonants.

The phonological items chosen are 4 vowels and 5 consonants, which are found to be projecting features that can relate to the Malay language spoken in Sabah. These vowels and consonants are termed variables for the purpose of this study. The phonological variables are (α), (↔), (ε), (ο), (η), (κ), (ρ), (τΣ) and (?). Each variable has 'variants'. For example, the variable (α) has two variants: [α] variant and [↔] variant.

All these linguistic phonological variables are enclosed in parentheses, as practised by the dialectologists (Labov 1966, Chamber and Trudgill 1980, Francis 1983, Wong 1985, Idris Aman 1995). Most of these variables are not identical to phonemes, since they examined in so far as their occurrence in certain positions only (see further explanation in 3.4.3.1-4, 6-9). However, there is some variables that are synonymous with phoneme (see 3.4.1.5).

Variants are then encoded with the variable in parentheses by an underscore number, for example variable (↔) with 7 variants, the coding will be from $(\leftrightarrow)_{-1}$ to $(\leftrightarrow)_{-7.}$ The first variant $(\leftrightarrow)_{-1}$ is the standard variant, while the others variants can be naturally arranged according to place of articulation. The greater the number of coding, the further is the variant from the standard variant, which is the STM. All comparisons are made with reference to STM.

3.4.3.1 VARIABLE (α)

In this study, the variable (α) represents solely the word-final /a/. However, phonological variations are not obvious in word-initial /a/ and word-medial /a/. There are two variants of the variable (α): $(\alpha)_{-1}$ = [↔] and $(\alpha)_{-2}$ = [α]. In other words, the word-final /a/ can be realised as mid central unrounded vowel [↔] and low back

unrounded vowel [α]. For example [ματα] ~ [ματ↔], [κατα] ~ [κατ↔] and [μανα] ~ [μαν↔]. These variant [α] and variant [↔] correspond to what Asmah Haji Omar (1988:138) termed as the 'α-variety', which the "a" is realised as [α], and the 'schwa-variety'[7], which the "a" is realised as [↔].

3.4.3.2 VARIABLE (↔)

The variable (↔) represents only the word-initial /↔/ and word-medial /↔/. This variable does not represent word-final /↔/, as it is only occurs in the preposition *ke* and affix *me.* There are five variants of the variable (↔) in initial position: $(↔)_{-1}$ = [↔], $(↔)_{-2}$ = [O], $(↔)_{-3}$ = [ε], $(↔)_{-4}$ = [ι] and $(↔)_{-5}$ = [α]. In other words, word-initial /↔/ can be realised as mid central unrounded vowel [↔], low back unrounded vowel [α], high front unrounded vowel [i], mid front unrounded vowel [ε] and ↔-deletion or ↔-dropping [O].

However, there are seven variants of the variable (↔) in medial-position: $(↔)_{-1}$ = [↔], $(↔)_{-2}$ = [O], $(↔)_{-3}$ = [ε], $(↔)_{-4}$ = [o], $(↔)_{-5}$ = [υ], $(↔)_{-6}$ = [ι] and $(↔)_{-7}$ = [α]. The word-medial /↔/ can be realised as mid central unrounded vowel [↔], low back unrounded vowel [α], high front unrounded vowel [i], mid front unrounded vowel [ε], mid back rounded vowel [o], high back rounded vowel [υ] and ↔-deletion or ↔-dropping [O]. For examples, [↔ναμ] ~ [αναμ], [λ↔βιη] ~ [λεβιη], [κ↔τΣιλ] ~ [κιτΣιλ], [s↔μυα]~ [συμυα] ~ [σομυα], and [β↔λασ] ~ [βλασ].

3.4.3.3 VARIABLE (ε)

The variable (ε) represents only word-initial /e/ and word-medial /e/. It does not represent word-final /e/. This is due to the reason that word-final /e/ only occurs in

[7] The 'α-variety' is based in the north, particularly in Alor Setar, while the 'schwa- variety' is based in the south, particular in Johor-Kuala Lumpur.

loan words such as *sate* and *tauge* (Nik Safiah Karim et.al. 1996). There are four variants of the variable (ε) in both word-initial and word-medial positions: $(\varepsilon)_{-1}$ = [ε], $(\varepsilon)_{-2}$ = [↔], $(\varepsilon)_{-3}$ = [α] and $(\varepsilon)_{-4}$ = [ι]. In other words, both word-initial /e/ and word-medial /e/ can be realised as mid front unrounded vowel [ε], mid central unrounded vowel [↔], low back unrounded vowel [α] and high front unrounded vowel [i]. Examples: [βολεη] ~ [βυλιη] and [τελεφον] ~ [ταλιπον] ~ [τ↔λεπον].

3.4.3.4 VARIABLE (o)

The variable (o) represents only word-initial /o/ and word-medial /o/. It does not represent word-final /o/. This is mainly because word-final /o/ only occurs in loan words such as *logo* and *polo* (Nik Safiah Karim et.al. 1996). There are two variants of the variable (o): $(o)_{-1}$ = [o] and $(o)_{-2}$ = [υ]. In other words, both the word-initial /o/ and word-medial /o/ can be realised as mid back rounded vowel [o] and high back rounded vowel [υ]. For example,[οραN] ~ [υραN], [rοσακ] ~ [rυσακ] and [κοτορ] ~ [κυτυρ].

3.4.3.5 VARIABLE (η)

The variable (η) is identical with the phoneme /h/. This variable (h) represents all word-initial /η/, word-medial /η/, and word-final /η/. The variable (h) has two variants in word-initial position: $(\eta)_{-1}$ = [h] and $(\eta)_{-2}$ = [O]. This word-initial /η/ can only be realised as pharyngeal fricative [h], and h-deletion or h-dropping [O].

However, variable (h) has three variants in word-medial and final positions: $(\eta)_{-1}$ = [h], $(\eta)_{-2}$ = [?] and $(\eta)_{-3}$ = [O]. Both word-medial /h/ and word-final /h/ can be realised as glottal fricative [h], glottal stop [?] and h-deletion or h-dropping [O]. For

example [ηιδυN] ~ [ιδυN], [παηα] ~ [πα?α] ~ [παα?] and [λιδαη] ~ [λιδα?]~ [λιδα].

3.4.3.6 VARIABLE (κ)

In this study, the variable (κ) represents solely the word-final /κ/. There are three variants of the variable (κ): $(κ)_{-1}$ = [κ], $(κ)_{-2}$ = [?], and $(κ)_{-3}$ = [O]. This word-final /κ/ can be realised as velar stop [κ], glottal stop [?], and k-deletion or k-dropping [O]. Examples: [ποκοκ] ~ [ποκο?] ~ [ποκο] and [τιδακ] ~ [τιδα?] ~ [τιδα].

3.4.3.7 VARIABLE (ρ)

The variable (ρ) represents solely the word-final /r/. This variable does not represent word-initial /r/ and word-medial/r/ since phonological variations are not obvious. There are three variants of the variable (ρ): $(ρ)_{-1}$ = [ρ], $(ρ)_{-2}$ = [λ], and $(ρ)_{-3}$ = [O]. In other words, the word-final /r/ can be realised as alveolar trill [ρ], alveolar lateral [λ], and r-deletion or r-dropping [O]. For example [βασαρ] ~ [β↔σαλ] ~ [β↔σα] and [κελυαρ] ~ [κελυαλ] ~ [κελυα]. The variants [ρ] and [O] suit Asmah's conception of 'the a-variety', which the "r" is realised as the flapped [ρ] and 'the schwa-variety', which the "r" is deleted [O] (Asmah Haji Omar 1988:138).

3.4.3.8 VARIABLE (τΣ)

The variable (τΣ) represents only the word-initial/τΣ/ and word-medial /τΣ/, since /τΣ/ does not occur in word-final position. There are two variants of the variable (τΣ): $(τΣ)_{-1}$ = [τΣ] and $(τΣ)_{-2}$ = [σ]. In other words, the word-initial /τΣ/ and word-medial /τΣ/ can

be realised as alveo-palatal affricate [τΣ] and alveolar fricative [s]. For example, [τΣαρι] ~ [σαρι] and [ματΣαμ] ~ [μασαμ].

3.4.3.9 VARIABLE (?)

In this study, the variable (?) represents exclusively the insertion of a glottal in word-final open syllables. However, this excludes single syllable words such as the preposition *δι* and *κε.* The variable (?) is considered as an insertion of glottal with reference to STM that the word-final open syllables are remain opened with a vowel. There are three variants of the variable (?): $(?)_{-1}$ = [O]; $(?)_{-2}$ = [η] and $(?)_{-3}$ = [?]. The variable (?) can be realised as glottal stop ?-insertion [?]; glottal fricative h-insertion [η]; and no glottal insertion [O] in word-final open syllables. For example, [ματα?] ~ [ματαη] ~ [ματα] and [απα?] ~ [απαη] ~ [απα].

3.5 CONCLUSION

The methodology of selecting focal areas, informants and variables has been discussed in this chapter. In all the three areas considerations are given to existing features. Other works have been referred to and adapted to suit the purpose.