The Role of Demographic Information in Understanding the Spatial Distribution of Multiple Language in Malaysia

Tarmiji Masron*, Usman Haji Yaakob* and FUJIMAKI Masami**

Abstract

The Population and Housing Census conducted every 10 years by the Department of Statistics, Malaysia is able to provide basic data for a variety of fields, among them is the geolinguistic field or language mapping. Basic information related to population such as total population, age, race, occupation, language and education may be obtained from this effort. With reference to Population and Housing Census information, this paper attempts to highlight and discuss how the data and demographic information can become the basic data in mapping and understanding the spatial distribution of multiple languages for geolinguistic study.

Ⅰ. Introduction

Geolinguistics refers to the study or science of linguistics in relation to geography[1]. Among other studies involving the mapping of language either in a large scale, such as political boundaries, administrative boundaries, boundaries of language (speaking), or more specifically refers to the individual speaking a language or dialect in the area. However, studies involving language mapping are too costly and the sampling method may be unclear especially in showing the distribution of a language/dialect in an area. In Malaysia, the population and housing census is carried out once every ten years. Basic information related to population such as total population, age, race, occupation, and education can be obtained from this effort. This paper attempts to highlight and discuss how the data and demographic information may be used as basic data in mapping and understanding the spatial distribution of multiple languages for geolinguistic study.

Ⅱ. Population Census and Demographic Information

In Malaysia, the first comprehensive census was held in 1891. The states participating in the census were the Straits Settlements, i.e. Singapore, Penang and Malacca and also the Malay States which were under the British protection, namely Pahang, Perak, Negeri Sembilan and Selangor[2]. The following censuses were carried out in 1901 and 1911. The first 'Pan-Malayan' census or also known as the

*　Section of Geography, School of Humanities, Universiti Sains Malaysia

** College of Letters, Ritsumeikan University

Key words : Demographic Information, Multiple Languages, Geolinguistics

'British Malay census' was held in 1921[3], which covered all states including the Straits Settlements (Labuan, Penang and Malacca), the Federated Malay States (Perak, Selangor, Negeri Sembilan and Pahang) and the Non-Federated Malay States (Johor, Kedah, Kelantan, Terengganu, Perlis, Sabah and Brunei). The census which used the de facto approach also includes questions on the scope of language and the ability to speak English.

The second and third Pan-Malayan censuses were held in 1931 and 1947 respectively. The last census conducted by the British colonial government on 17–18th of June 1957 covered the states in the Federation of Malaya. According to Sidhu (1975)[4] the previous censuses are important sources that can be used by researchers from various disciplines who wish to study various aspects of population in Peninsular Malaysia. For Sabah which has conducted multiple limited censuses since 1901 and 1911, and Sarawak which held the first population census in 1939, both states have a lot of socio-economic and demographic data that can be benefitted in

Table 1 Questions Topics by Year of Census in Malaysia

Question Topics	1947	1951	1957	1960	1970	1980	1991	2000
De facto	X	X	X	X	X	X		
Residence								
Place of birth	X	X	X	X	X	X	X	X
Duration of residence					X	X		
Previous residence					X	X		
Urban/Rural	X		X	X	X	X	X	X
Gender	X	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X	X
Relationship with head of household	X		X	X	X	X	X	X
Marital status	X	X	X	X	X	X	X	X
Number of children born alive	X	X	X	X	X	X	X	X
Children living with parent	X	X	X	X	X	X	X	X
Nationality			X		X	X	X	X
Literate	X	X	X	X	X	X	X	X
School enrolment			X		X	X	X	X
School attendance			X	X	X	X		
Education level					X	X	X	X
Ethnic/racial origin	X	X	X	X	X	X	X	X
Language					**X**	**X**		
Religion	X	X			X	X		
Household composition								
Economic activity	X	X	X	X	X	X	X	X
Occupation	X	X	X	X	X	X	X	X
Industry	X	X	X	X	X	X	X	X
Employment status	X		X	X	X	X	X	X
Income								
Housing	X	X		X	X	X	X	X

Source: Adapted from Domschke, E and Goyer, D. S.: *The Handbook of National Population Censuses—Africa and Asia*, Greenwood Press, New York, 1986, pp. 745–758.

the situation before and after they joined the Federation of Malaysia alliances in 1963. As a sovereign country, Malaysia's first population census was conducted in 1970 and followed in 1980, 1991 and 2000. Reviews on the characteristics and critical issues of population census of Malaysia 2000 have been conducted[5].

The requirements in population census information gathering vary according to local and national interests. These include the concepts and topics of questions that must be prepared in the census, and it should be in accord with international requirements to analyze the population profile or stock between countries and make comparisons, such as growth rate and age group and gender, socio-economy, urbanization level, population employment according to categories of industry and space. List of census questions topics in Malaysia since 1947, 1957, 1970, 1980, 1991 and 2000 are shown in Table 1.

Ⅲ. Demographic Information and Language Mapping

From Table 1, although there are changes in the concentration of question or list of questions being asked during the process of collecting census data, particularly those involving language, but generally population census provides population data, among others, referring to a race or ethnic group. Language (in daily speech) is also generally refers to a race or ethnic group.
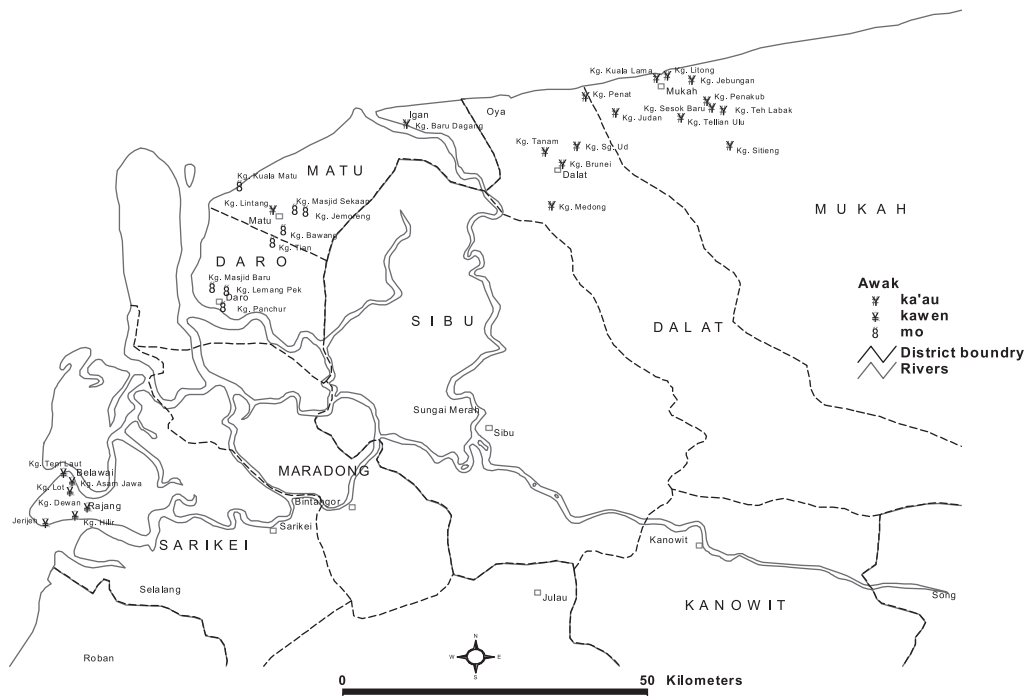

Figure 1 Variation of the Word 'You' Used in Majority by the Respondents

Source: Salasiah Che Lah, Tarmiji Masron, Azimah Sazali, Ruslan Rainis, Suraiti Ghazali, Alias Abd Ghani & Salfarina Abdul Ghafor: The Spatial Lexical Variation and Distribution of Melanau Language, *Sarawak Museum Journal* LXIII (84), 2007, pp. 9941–9964.

Therefore, the available census information can be used as a basis to develop or create a language map for the country, state, district, county or census block in Malaysia.

There are several studies and writings found in the country and abroad that used information from the census[6]. At the national level, there is no study and writing that tries to see the role of demographic information or statistics from the department of statistics as a source in language study especially in language mapping.

In the aspect of understanding language in terms of space and language mapping, there also are many studies[7]. Study by Salasiah et al. (2007)[8] has tried to identify the spatial distribution of Melanau dialect in the settlements of Melanau people in Sarawak. Figure 1 showed one of the language map generated. This study involved mapping at the individual respondent level.

However, in Malaysia, no efforts have been made to look at the role that can be played by population data through the national population and housing census as a data source to develop a language map.

1. Population and Spoken Language

From the population and housing census carried out, there are two census years that have topic directly related to language which is the 1970 and 1980 censuses. According to the General Report of the Malaysian Population, language-related information was issued only for Peninsular Malaysia (Table 2), while for the Year 1980 Report data relating to language covers Peninsular Malaysia, Sabah and Sarawak (Table 3). Information from both census years can be used by geolinguist to understand the language scenario in Malaysia. However, the report released by Department of Statistics to the public only shows information at the national scale, and does not refer to state or district or county or census block. Nevertheless, since the census involved individuals and census blocks, it is believed that the language information referring to the census block or county or district as well as the state from the two census years can be obtained.

From Table 2, the community of a society influenced the use of their language and their proficiency level to Malay language. From the table, Malay language, Mandarin and English were the main languages used or have a high level of proficiency for the population in Peninsular Malaysia. The Malay Language proficiency level (fluency) among the Chinese and Indian community is relatively low. However, they were able to communicate using their fundamental Malay language.

From Table 3 in 1980, Bahasa Malaysia, English as well as Chinese language were the main spoken languages used by urban residents in Malaysia and the rest were languages such as Tamil, Iban, Kadazan/Dusun, and other languages. Almost similar scenario occurred in rural areas with Bahasa Malaysia, English and Chinese language, were the main spoken languages by the population in Malaysia. This data is directly related to the total population in Malaysia, whether they are urban or rural population. The pattern of language used by the people in Malaysia would be more apparent if the data at the state, district, county and census block level is used. Furthermore, a language map can also be developed if data at the above-mentioned scale is used.

Table 2 Languages Spoken by Population Aged 10 Years and above: Peninsular Malaysia

Spoken Language	Malay			Chinese			Indian			Others			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
Malay (Fluent)	115,827	108,149	223,976	24,958	17,239	42,197	19,622	11,170	30,792	1,599	1,449	3,048	162,006	138,007	300,013
Malay (Fundamental)	2,511	2,283	4,794	128,084	98,333	226,417	39,343	28,151	67,494	2,035	2,015	4,050	171,973	130,973	302,755
English	41,753	24,457	66,210	121,704	94,753	216,457	46,579	29,742	76,321	7,197	7,311	14,508	217,233	156,263	373,496
Mandarin	435	244	679	132,977	113,663	246,640	281	153	434	173	161	334	133,886	114,221	248,087
Tamil	2,335	1,288	3,623	557	363	920	77,763	58,942	136,705	344	282	626	80,999	60,875	141,874
*Other languages	3,962	2,611	6,573	279,994	302,834	582,828	13,124	10,225	23,379	2,799	2,828	5,627	299,879	318,528	618,407
+Total	119,391	111,383	230,774	330,326	350,041	680,367	92,517	70,993	163,510	7,783	8,027	15,810	550,017	540,444	1,090,461

* 'Other languages' includes other languages for the Chinese and Indians apart from Mandarin and Tamil.
+ The figures were double counted and will not make the total.
Source: Department of Statistics Malaysia: *Laporan AM, Banci Penduduk Malaysia 1970, Vol. 2*, Kuala Lumpur, 1975, pp. 153–155.

Table 3 Population Aged 10 and over by the Spoken Language, Gender, and Strata, Malaysia: 1980

Spoken language	Urban			Rural			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Fluent Malay*	894,160	832,135	1,726,295	1,859,984	1,815,593	3,675,577	2,754,144	2,647,728	5,401,872
Fundamental Malay*	587,234	544,623	1,131,857	852,235	791,600	1,643,835	1,439,469	1,336,223	2,775,692
English	767,701	616,493	1,384,194	515,166	356,649	871,815	1,282,867	973,142	2,256,009
Chinese*	824,407	863,626	1,688,033	690,656	668,870	1,359,526	1,515,063	1,532,496	3,047,559
Tamil	164,250	146,224	310,474	226,270	221,272	447,542	390,520	3,674,496	758,016
Iban	15,825	11,237	27,062	163,224	152,219	315,443	179,049	163,456	342,505
Kadazan/Dusun	5,259	4,518	9,777	82,004	81,662	163,666	87,263	86,180	173,443
Bajau	7,212	5,816	13,028	42,848	42,202	86,050	51,060	48,018	99,078
Other languages	47,442	42,263	89,705	196,772	185,281	382,053	244,214	227,544	471,758
Mute	1,712	1,679	3,391	4,890	4,459	9,349	6,602	6,138	12,740
Total+	1,672,019	1,698,270	3,370,289	3,043,495	3,097,888	6,141,383	4,715,514	4,796,158	9,511,672

* For definition, see "notes and list of terminology".
+ Figures were double counted and did not make the total.
Source: Department of Statistics Malaysia: *Genaral Report of the Population Cancus, Vol. 2*, Kuala Lumpur, 1983, p. 76.

2. Racial Segregation and Language

Segregation is a process or an outcome. It is usually seen in five dimensions, namely equality, exposure, concentration, centralization and clustering[9]. Study of segregation by race and class has been initiated by sociologist and was an old theme since the 1920s in this field. Segregation studiy is important in the effort to create a balanced society either in terms of features (structure) or space which involves government policy[10].

The most prevalent and popular segregation measure, especially among geographers and sociologists is the dissimilarity index, *D*, or the segregation index that was introduced by Duncan and Duncan (1955)[11]. However, this method can only take into account two groups of people such as white and black. One of the segregation calculation methods for multiple groups that can calculate each spatial subunit and is suitable for the purpose of mapping or spatial analysis using GIS is the entropy index.

It is a common statistical measure for distribution uniformity[12]. Generally in Malaysia spoken language has a direct relationship with a race, this index can be used as proxies to look at equality, exposure, concentration, centralization or clustering of a race and eventually language in spatial context.

With reference to Penang, Figure 2 clearly shows the spatial distribution pattern and the distribution of racial segregation phenomenon in the state. In general, moderate segregation is the common segregation category found in this state with 28 Mukim in this category. It is followed by 26 moderately low segregation, 22 moderately high segregation, while the remainder are 5 highly segregated and 3 in low segregation[13]. These spatial pattern provides general picture to language researcher, for example, 5 mukim that belong in high segregation category showed domincance of a particular race in the counties, where indirectly their spoken language is just about the same or they
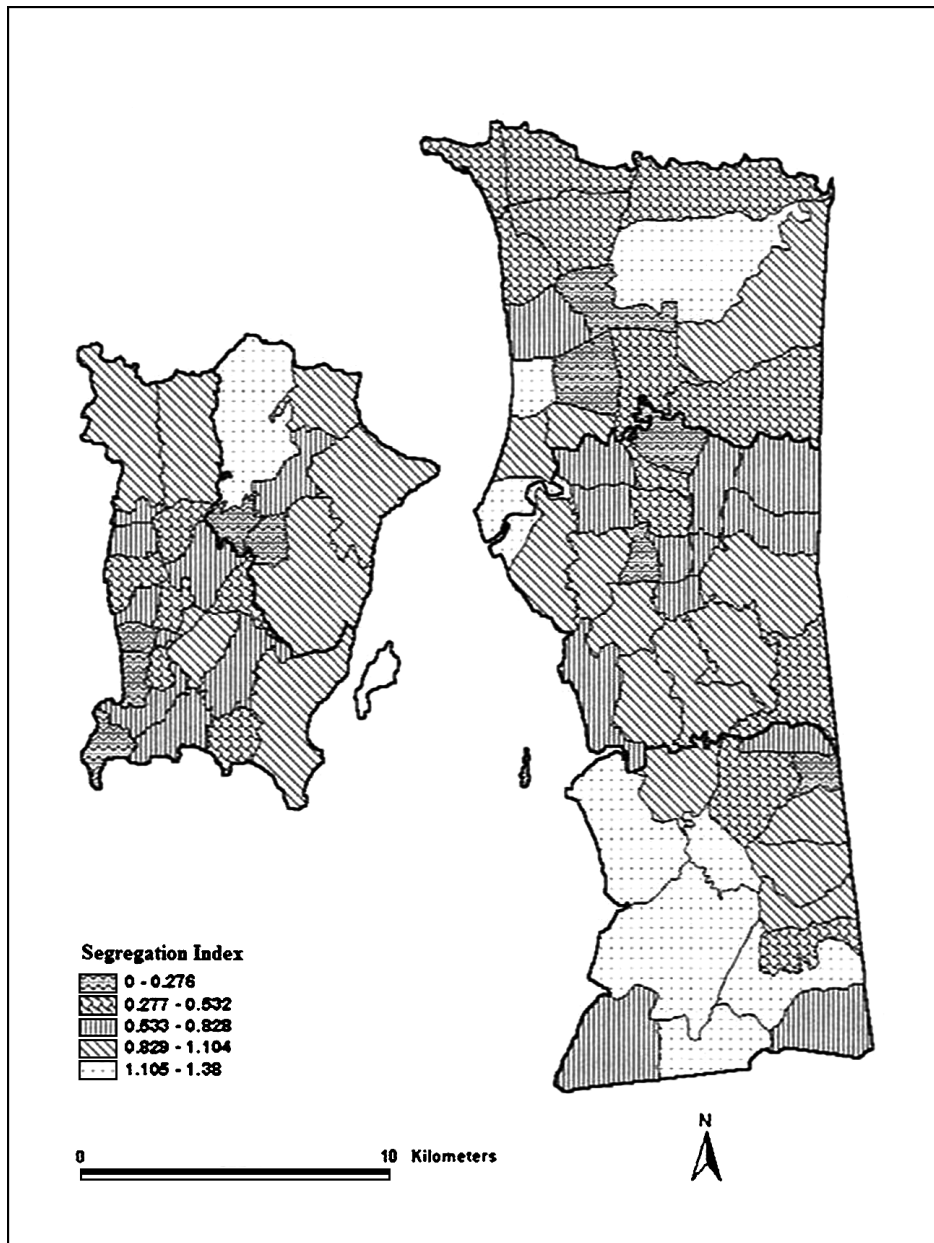
Figure 2 Distribution Pattern of Racial Segregation in Penang: 1991

Source: Ruslan Rainis dan Tarmiji Masron: Corak Ruangan Segregasi Kaum di Negeri Pulau Pinang Pada Tahun 1991, *Jurnal Ilmu Kemanusiaan* 9, 2002, pp. 1–25.

only use one language.

3. Socio-Economic Information

Socio-economic information is among the important information that can be found from the Population and Housing Census Report. This information, among others, contains data on education or the educational background of inhabitants in a census area, whether it is the state, district, county or census block. Population census data relating to education was commenced to publish since 1970 and has been continuously shown every census year. The data was about the population education level according to their highest education attainment and age group of 0–4 to 65 years and older. In reference to Sarawak based on the 1970 population census information it was found that 60.4 per cent of the population did not go to school, 30.4 percent went up to primary school, 6.1 percent up to lower secondary school and 3.4 percent finished their upper secondary school (Table 4). No information about tertiary education was published. Information in this table also showed that more than 80.0 percent of the population aged 45–49 and above did not go to school. Apart from the population in the age group 25–29 up to 40–44 also more than 50.0 per cent did not go to school. Population in the age group that was supposed to be in primary school which is the 5–9 years group, only 45.9 percent went to school, followed by 10–14 years with 70.3 per cent and 15–19 years group with 40.9 percent.

Table 4 Population Distribution by Highest Education Attainment in Sarawak: 1970

Age Group	Total (Number)	Not Schooling (%)	Primary School (%)	Lower Secondary School (%)	Upper Secondary School (%)	Total (%)
0–4	149,629	100.0	—	—	—	100.0
5–9	149,243	54.1	45.9	—	—	100.0
10–14	107,594	18.2	70.3	11.3	0.2	100.0
15–19	89,998	27.2	40.9	20.1	11.8	100.0
20–24	70,305	42.0	35.6	12.3	10.1	100.0
25–29	58,985	54.2	30.6	9.1	6.1	100.0
30–34	49,491	63.1	25.8	6.9	4.2	100.0
35–39	44,692	71.1	21.4	4.7	2.8	100.0
40–44	38,932	76.1	18.5	3.4	2.0	100.0
45–49	32,585	80.0	15.1	3.2	1.7	100.0
50–54	30,667	83.0	12.8	2.6	1.6	100.0
55–59	20,726	81.2	15.1	2.2	1.5	100.0
60–64	19,279	86.1	11.3	1.6	1.0	100.0
65+	25,166	87.5	10.4	1.3	0.8	100.0
TOTAL	887,292	535,602 (60.4%)	270,300 (30.4%)	54,006 (6.1%)	27,384 (3.1%)	100.0%

Source: Department of Statistics Malaysia: *Jadual-jadual Asas Penduduk Sarawak, Part XIII, Jilid 1*, Kuala Lumpur, 1976, pp. 153–155.

Overall, the total Sarawak population that went to primary school was 30.4 percent. This was followed by lower and upper secondary school, 6.1 per cent and 3.1 per cent respectively. Comparing the achievement of 1970 and 1980 showed little changes especially at the secondary level. For example, there was still 72.0 per cent population aged 45–49 years did not go to school followed by 77.0 per cent of 50–54 years and for the next age group with more than 80.0 percent. Nevertheless admission to lower and upper secondary was rising, particularly for the age groups 10–14 and 15–19 years. If in 1970, admission to upper secondary was less than 1.0 percent but in 1980 the number has increased to six per cent. This performance showed the awareness of Sarawakians to attain higher education and this development can be observed in the 1991 and 2000 population census. These informations relating to education can be used to observe, especially the population exposure to the official languages for example Bahasa Malaysia and English as well as other languages proficiency that can be obtained from school.

Ⅳ. Conclusion

The Population and Housing Census collected every 10 years is the best source to investigate information relating to population and language in Malaysia. Efforts to gather language or dialect information separately requires a high cost. Therefore to create a data bank relating to population and language in Malaysia can later lead to construction of a language map or geolinguistic analysis at the census block, county, district, state as well as national level. The data bank and language map are important especially to policy makers as well as researchers in observing and assessing the language scenario whether from the aspects of language proficiency, whole language or dialect, daily spoken language and many more. In Malaysia, the daily spoken language is still greatly related to its race. Therefore, for the time being, information related to race as well as socio economic information can still be used as proxies in developing language map for the population in Malaysia. It is hoped that the Population and Housing Census in future will take into consideration information relating to language and continues to do it consistently.

Notes

1) http://www.thefreedictionary.com/geolinguistics (Accessed on 17 April 2011)
2) Domschke, E. and Goyer, D. S.: *The Handbook of National Population Censuses*, Greenwood Press, New York, 1986, p. 746.
3) ibid. 2), p. 747.
4) Sidhu, J. S.: Problems of Census-Taking in Peninsular, *Ilmu Alam*, 4 (Mei), 1975, pp. 29–44.
5) Usman Hj. Yaakob: The Malaysian Census 2000; Characteristics and Critical Issues, *Journal of the Malaysian Branch of the Royal Asiatic Society*, LXX1X (Part 1), 2006, pp. 27–42.
6) ① Smith, T. E.: *Population Growth in Malaya*, Royal Institute of International Affairs, London, 1951. ② Lee, Y. L.: Population Changes in Sabah, 1951–60, *The Journal of Tropical Geography* 26, 1968, pp. 55–67. ③ Hamzah Sendut: Urbanization in West Malaysia 1911–70, *Third International Conference Modernization of Asia*, Universiti Sains Malaysia, Penang, September 3–8, 1972. ④ Usman Hj. Yaakob: *Prinsip Kajian Kependudukan*, Dewan Bahasa dan Pustaka, Kuala Lumpur, 1989. ⑤ Kok, K. L.: Components of Ethnic Urban Population Growth, *Ilmu Masyarakat*, (Julai-Disember), 1992, pp. 40–66. ⑥ Department of Statistics Malaysia: *Migrasi Dalaman Di Malaysia*, Siri Monograf Banci

Penduduk No. 2, Kuala Lumpur, 1996. ⑦ Usman Haji Yaakob: Perubahan Penduduk dan Perbandaran di Malaysia, *Kertas Kerja Persidangan Kebangsaan Geografi, Anjuran Jabatan Geografi*, Universiti Malaya, Kuala Lumpur, 15–17 Mei 2001. ⑧ Department of Statistics Malaysia: *Migrasi Dan Taburan Penduduk*, Kuala Lumpur, 2004. ⑨ Usman Hj. Yaakob: *Kesihatan dan Kematian Bayi di Kedah*, Penerbit Universiti Sains Malaysia, Pulau Pinang, 2008.

7) ① Collins, J. T.: *Dialek Ulu Terengganu*, Penerbit Universiti Kebangsaan Malaysia, Bangi, 1983. ② Collins, J. T. and Voorhoeve, C. L.: Moluccas (Maluku), in S. A. Wurm and S. Hattori eds.: *Language Altlas of the Pacific Area, 2* (Lembar 45), Australian Academy of the Humanities in Collaboration with the Japan Academy, Canberra, 1983. ③ Rohani Mohd Yusof: *Dialek Geografi di Kuala Kangsar; Satu kajian Perbandingan Fonologi*, Tesis Sarjana, Universiti Malaya, Kuala Lumpur, 1996 (unpublished). ④ Collins, J. T.: Bahasa Melayu di Sungai Mahakam; Siasatan Fornologi Dialek Kutai, *Sari* 10, 1991, pp. 3–24. ⑤ Domschke and Goyer: *The Handbook of National Population Censuses, Africa and Asia*, Green Wood Press, New York, 1986, pp. 754–755. ⑥ Patriantoro dan Sudarsono: *Geografi Dialek Bahasa Melayu di Kabupaten Sanggau* (unpublished). ⑦ Rivero, A., Llull, G. and Merlo, G. D.: Mapping the Spatial Distribution of Language, Universidad Nacional de San Juan, Argentina, 2002. http://www.esri.com/news/arcuser/1002/linguistics.html (Accessed on 5 May 2004). ⑧ Boberg, C.: Geolinguistic Diffusion and the U.S.—Canada Border, *Language Variation and Change* 12, 2000, pp. 1–24. ⑨ Anderbeck, Karl R.: *Malay Dialects of the Batanghari Rivers Basin (Jambi Barat)*, Graduate Thesis, Universiti Kebangsaan Malaysia, Bangi, 2003 (unpublished). ⑩ Salasiah Che Lah et al.: *On Defining the Model of Spatial Distribution and Diffusion of Melanau Language Family along Rejang-Baram Rivers*, Laporan Akhir Geran Penyelidikan (FRGS), 2004 (unpublished). ⑪ Asmah Haji Omar: Regional Varieties of Malay in *the Encyclopedia of Malaysia*, *Languages and Literatures* 9, 2004, pp. 26–27. ⑫ Collins, J. T.: *Dialek Melayu di Sungai Mahakam: Pemetaan dan Deskripsi*, Pusat Pengurusan Penyelidikan, Universiti Kebangsaan Malaysia, Bangi, 2004. ⑬ Collins,

J. T.: Dialek Melayu di Sungai Mahakam: Pemetaan dan Deskripsi (The Malay Dialects of the Mahakam Rivers Basin: Mapping and Description), in Rahimah Abdul Aziz ed.: *Sari Penyelidikan UKM. 204. UKM Research Highlights*, Universiti Kebangsaan Malaysia, Bangi, 2005, pp. 115–121. ⑭ Chong Shin: Menilai Distribusi dan Ciri Linguistik Ibanik, in Chong Shin, Karim Harun and Yabit Alas eds.: *Reflections in Borneo Rivers; Essays in Honour of James T. Collins (Book 1)*, STAIN Pontianak Press, Pontianak, 2006, pp. 97–109.

8) Salasiah Che Lah, Tarmiji Masron, Azimah Sazali, Ruslan Rainis, Suraiti Ghazali, Alias Abd Ghani and Salfarina Abdul Ghafor: The Spatial Lexical Variation and Distribution of Melanau Language, *Sarawak Museum Journal* LXIII (84), 2007, pp. 9941–9964. [Refer to the following reports; ① Tarmiji Masron, Ruslan Rainis, Suraiti Ghazali, Salasiah Che Lah, Azimah Sazali, Alias Abd Ghani and Salfarina Abdul Ghafor: Using GIS to Grasp Dialectal Variation, Map Asia 2005, Hotel Senayan, Jakarta. ② Tarmiji Masron, Ruslan Rainis, Suraiti Ghazali, Salasiah Che Lah, Azimah Sazali, Alias Abd Ghani and Salfarina Abdul Ghafor: Mapping the Spatial Lexical Variation and Distribution of Melanau Language, the 9th International Geographical Conference in Taiwan, November 11–14, 2005]

9) Massey, D. S. and Denton, N. A.: The Dimensions of Residential Segregation, *Social Forces* 67, 1988, pp. 281–315.

10) Ruslan Rainis dan Tarmiji Masron: Corak Ruangan Segregasi Kaum di Negeri Pulau Pinang Pada Tahun 1991, *Jurnal Ilmu Kemanusiaan* 9, 2002, pp. 1–25.

11) Duncan, D. and Duncan, B: A Methodological Analysis of Segregation Indexes, *American Sociological Review*, 20, 1955, pp. 210–217.

12) ibid. 10.

13) ibid. 10.

マレーシアにおける多言語空間分布を理解する際の
人口統計情報の役割

タールミジィ・マスロン＊、**ウスマン・ハジ・ヤァクブ**＊、**藤巻正己**＊＊

　マレーシア統計局によって 10 年ごとに実施されてきた人口・住宅センサスは、地理―言語学の分野あるいは言語地図の作成など、さまざまな分野に対して基本データを提供することができる。すなわち総人口、年齢、種族、職業、言語、教育といった人口に関連する基本情報は、人口・住宅センサスから入手できる。本稿は、人口・住宅センサス情報にかかわって、これらのデータや人口統計情報が地理 - 言語学的研究にとって、多言語空間分布のマッピングや理解に際して、どのように基礎的データとなりうるかについて関心を払い、検討を加えるものである。

キーワード：人口統計情報、多言語、地理―言語学

＊　マレーシア科学大学人文学院地理学系
＊＊ 立命館大学文学部