Geo-linguistics study on lexical and phonology dialect variations in North Perak, Malaysia

Harishon Radzi a, Nor Hashimah Jalaluddin b, Zaharani Ahmad c, Siti Noraini Hamzah d, Yusmaniza Mohd.Yusoff e & Norlisafina Sanit f.

School of Language Studies and Linguistics, Faculty of Social Science and Humanities
Universiti Kebangsaan Malaysia 43600 Bangi, Selangor, MALAYSIA

Abstract

Most of the geographical dialect studies in Malaysia use a traditional, manually drawn map in determining dialect boundaries or isoglosses. This impression-based drawing method is scientifically not appropriate because the isoglosses produced are unclear and ambiguous. Using current modern technology with an interdisciplinary approach such as geo-linguistics has led to a breakthrough as a remedy for this problem. A Geographical Information System (GIS) is more reliable in constructing an accurate line of isogloss for lexical variations and spatial distribution. Thus this paper aims to provide a better map presentation of lexical and phonology dialect variations found in North Perak, Malaysia. The data analysed are ‘mereka’ (they) and ‘ular’ (snake) respectively.

Keywords: northern Perak; dialect; GIS; lexical variation; lexical distribution; phonology.

1. Introduction

Dialect geography is the study of variations with regard to the geographic distribution of dialects, as well as how their distribution may be affected by geographical factors such as a mountain range, river basin, forest belt, etc. The regional variation is subsequently displayed in the form of a linguistic map wherein each dialect is demarcated by a linguistic boundary known as an isogloss. All geographical dialect studies in Malaysia have utilized a traditional manually drawn map in demarcating the isoglosses. For instance, based on phonological and lexical linguistic features, Ajid (1985) illustrates the isoglosses of local sub-dialects of Kelantan Malay spoken in Pasir Mas, Kelantan, while Rohani (1986) exhibits the isoglosses of Malay dialects spoken in Kuala Kangsar,

*Corresponding author.
E-mail address: naslin@ukm.my

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of Universiti Kebangsaan Malaysia.

Keywords: northern Perak; dialect; GIS; lexical variation; lexical distribution; phonology.
Perak. This traditional impressionistic-based linguistic mapping method is scientifically not reliable because the isoglosses drawn are not accurate and dubious.

In accounting for this shortcoming, recent developments in a dialectological study has opted for an interdisciplinary approach that is supported by a modern technological tool such as Geographical Information System (GIS). This recent study is known as geo-linguistics and is the blending of geographical elements into linguistics especially with regard to dialects. Special software known as Geographical Information System (GIS) is integrated in the dialect study. The GIS tool employs an interpolation technique that can store large volumes of spatial data, perform analysis and produce cartographic map results that are more reliable in determining the isoglosses of dialectal variations and spatial distribution. Teerarojanarat & Tingsabadh (2011) have demonstrated that the GIS successfully manages to produce reliable dialectal boundaries for central and non-central Thai based dialects on 170 semantic units. By utilizing the same GIS technique and approach, this study attempts to demonstrate the lexical variation ‘mereka’ (they) and phonological variation ‘ular’ (snake) distribution in northern Perak, Malaysia.

2. The Past and Present Of Dialect Research

Many research studies have been conducted on Malay dialects. For example, the research studies on Malay dialects focus more on a certain area and its distinctive features. Ajid (1985), Rohani (1986) and Collins (1983) have written on the geographical dialects of Pasir Mas, Kuala Kangsar and Ulu Trengganu respectively. Each of these studies contain phonological descriptions in determining the isogloss of each of the particular areas. Nevertheless, the description given is merely based on linguistics analysis without taking geographical information into account.

Researchers from Thailand are very involved in dialect and GIS. Premsrirat et. al (2004) reported the use of GIS as a geographical database to store and map the distribution of ethnolinguistic groups for the entire country. Meanwhile, Teerarojanarat and Tingsabadh (2008) have developed a geographical database that stores 170 semantic units in a study of lexical variation covering the entirety of Thailand and using the benefit of map display to reveal word distributions. In addition, Teerarojanarat and Tingsabadh (2011) have further explored central and non-central dialects using the overlay technique. The overlay technique integrates two types of data that leads to the systematic isogloss of the particular dialect concerned. It is the right time for Malay dialectologists to embrace a geolinguistics, multidisciplinary approach to better account for Malay dialects.

2.1 Focus of Study

This study focused on the use of Malay lexicals, [mereka] and [ular], in Northern Perak especially in Larut Matang and Selama (LMS) districts and Hulu Perak (HP) district. We attempted to demonstrate the entire process from data collection, data analysis and reporting of findings. Evidently, from the data gathered we found both lexical variations and phonological distributions in northern Perak. For instance, [ular] ‘snake’, is pronounced in three variants, namely [ula], [uls] dan [ulaf]. Apparently these variants served as a linguistic identity of the speakers. Our early assumption was that factors of boundaries, highlands, lowlands and demographics such as migration and transitions of ages played an important role in determining the dialect spoken. Simultaneously, economic factors had an impact as well.

The following geographical information was very useful in providing a background of northern Perak namely Larut Matang and Selama and Hulu Perak:

- Bordered by Kedah (to the west), southern Thailand (to the north) and Kelantan (to the northeastern).
- An array of highlands dissected the zone into two. One side faced Kedah and the other faced southern Thailand and Kelantan.
- A big river known as Sungai Perak originates from Hulu Perak and ends at a southern zone (Lenggong and Kuala Kangsar) where the lowlands laid.
A large lake to the east of Hulu Perak forms a large area of wetlands and is very lightly populated. Most of the population in this area were Aslian people.

A large area of northern Perak especially to the west has been greatly influenced by northern dialect. This has been proven by Asmah (1993). The sub-districts are Taiping, Bagan Serai, Selama and Bukit Gantang. However, the area that forms Lenggong is influenced by the Perak dialect in the south and Malay Pattani and Kelantan in the north and east.

3. Methodology

Since we were dealt with respondents directly in the field, proper planning to capture the reliable sources was very crucial. It was necessary that we had a key respondent from each selected sub-district and village to assist the research team. A preliminary survey was conducted. We approached the sub district head known as Penghulu to determine the Malay native speakers in the area. The Penghulu subsequently contacted the village head in order to identify the speakers. It is already known that migration patterns among people in the Malay archipelago are fluid, and as a consequence we can find many sub-ethnic Malays such as Rawa, Kerinci, Jawa, Banjar who speak differently from Malay. These people were excluded and were not selected as our respondents. However, migrants from Pattani were included since they have been in Malaya for over a millennium and they call themselves as Pattani Malay. They apparently speak Pattani Malay with influences from southern Thailand and Kelantan (eastern Malaysia). Their dialect is intelligible and well understood by their Malay neighbours.

After a proper screening, 38 villages were selected demographically. A purposive sampling ranging from four categories of age and gender were chosen from the native Malay speakers. Interviews and questionnaires were used as tools in gathering the relevant data. A group of 40 linguistics students equipped with phonetics knowledge were selected to carry out the interviews. They were accompanied by the village head to make sure that their safety was guaranteed. A list of 200 words or lexical items representing the most frequent Malay words were selected and interviewed. Apart from the wordlists, we collected the biodata and social background of the respondents and the topography of the village. It is apparent that topographic information plays a significant role in determining sub-dialects distribution. For instance, lexical variations of pronouns existed between 38 different villages. In addition to age gap, topographical features such as rivers and highlands seemed to play an important factor in dissecting the dialects. After the data was collected, the researchers transferred the information into the Microsoft Office Excel program. All the variants that have been identified were listed in a Microsoft Office Excel spreadsheet. Subsequently, the data was transferred again, this time into the GIS software.

Geographical Information System (GIS) is a software that incorporates the linguistic data together with its Spatial Analysis Tool. The spatial based technique uses a GIS tool to integrate the conventional linguistic approach and can help to produce a more systematic isogloss for the sub-dialects spoken in northern Perak. In this study, it is shown that new forms can be explained as language change, and it is also possible to see the spatial features of the areas where the new forms emerge and are disseminated.

4. Findings and Discussions

This study focused on the distribution of lexical and phonological variations in northern Perak. The lexical ‘mereka’ (they) was analysed for its variations distribution whereas the word ‘ular’ (snake) was analysed for its phonological variations distribution.

4.1 The Distribution of [mereka] ‘they’

‘Mereka’ is the third person pronoun in Malay language. The pronoun ‘mereka’ has nine variants. To reiterate, the combination of demography, topography and linguistics integrated with the GIS tool served as a systematic
window in understanding the distribution of this pronoun. As we can see in Table 1 below, three sets of the phonological variants are produced: i) \([\text{dema}]\), \([\text{dem}a]\) and \([\text{dem}a]\); ii) \([\text{depa}]\) and \([\text{dep}a]\); and iii) \([\text{hangpa}], \text{[mika]}, \text{[məyəkə]}\) and \([\text{moha}]\) which formed different lexical variants for ‘mereka’. As mentioned earlier, phonological variants existed in tandem with lexical variants, such as \([\text{dema}]\), \([\text{dem}a]\) and \([\text{dem}a]\). More interestingly the variants occurred at different places. Since these variants are phonemics, we considered them as three different variants.

In our analysis, \([\text{dema}]\), \([\text{depa}]\), \([\text{hangpa}]\) were more dominant in LMS. This distribution has to do with the geographical features mentioned above in 2.1. In the case of \([\text{dem}a]\), it is spoken in HP (Hulu Perak) with a score of 100%. HP is highly influenced by southern Thai and Kelantan which use the back semi low vowel /ɔ/. Unlike the pronoun \([\text{dem}a]\), it is equally distributed at 50% for each district. The reason for this distribution is that these places are located near the lowland where language contact occurs nearly every day. From the map computed by GIS tool, we saw that the frequent use of different variants was affected by demographic and geographic factors. This area is normally a densely populated area with many access roads to the neighbourhood area. Furthermore this area acts as a central point for all types of economic activities. More interestingly, the area has become a melting pot of different people with different sub-dialects who interact and communicate with each other. Consequently, the existence of different variants of the pronoun is well expected. Once again, the features of lowland geography and language contact have become the contributing factors in determining the distribution. Finally \([\text{mika}],[\text{məyəkə}]\) and \([\text{moha}]\) are only used in HP. These forms are either influenced by Perak dialect to the south or the Pattani Malay who came to this region over a century ago and mixed well with local communities. This area known as Lenggong is very closely related to Kuala Kangsar which has a very strong accent for Perak dialect.

<table>
<thead>
<tr>
<th>Lexical (L)</th>
<th>Larut Matang &amp; Selama</th>
<th>Hulu Perak</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1. dema</td>
<td>1 (25%)</td>
<td>0 (0%)</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>L2. dema</td>
<td>2 (50%)</td>
<td>2 (50%)</td>
<td>4 (50%)</td>
</tr>
<tr>
<td>L3. dema</td>
<td>1 (25%)</td>
<td>4 (100%)</td>
<td>5 (62.5%)</td>
</tr>
<tr>
<td>L4. depa</td>
<td>4 (100%)</td>
<td>2 (50%)</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>L5. depa</td>
<td>2 (50%)</td>
<td>1 (25%)</td>
<td>3 (37.5%)</td>
</tr>
<tr>
<td>L6. hangpa</td>
<td>2 (50%)</td>
<td>0 (0%)</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>L7. mika</td>
<td>0 (0%)</td>
<td>1 (25%)</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>L8. məyəkə</td>
<td>0 (0%)</td>
<td>1 (25%)</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>L9. moha</td>
<td>0 (0%)</td>
<td>1 (25%)</td>
<td>1 (12.5%)</td>
</tr>
</tbody>
</table>

The distribution of these variants is well represented in the map processed with GIS. In this map, colours play an important role in describing the distribution of the variants. From the map below, places with one or two variants are surrounded by highlands especially in Larut Matang and Selama (yellow colours). As a result, the dissemination of other variants to the area are obstructed. We were told by respondents from this area that they can hardly understand dialects spoken by people from the other side of the hill. They are entirely two different sub-dialects. Nevertheless, places with purple and blue show the most variants distributed. As mentioned earlier,
this is due to geographical factors such as lowlands and river basins. We can see this at Hulu Perak district (denoted by ♦). A few villages with 5 - 7 variants in Larut Matang and Selama as well were identified (denoted by ●). These factors lead to economic activities and simultaneously attract people to meet and create language contact. In this area we can find a combination of either Kedah and Pattani dialect or Perak, Kedah and Pattani dialect.

4.2 The Distribution of [ular] ‘snake’

The lexical [ular] has its own variants. We identified 3 variants for [ular] ‘snake’, namely [ula], [ulɔ] and [ulaf]. These variants are influenced by well-distributed patterns and assimilated from the neighbouring area. Variant [ula] is pronounced dominantly in this area. The next dominant variant is [ulɔ], followed by the variant [ulaf] as shown in Table 2.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Larut Matang &amp; Selama</th>
<th>Hulu Perak</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1 - ula</td>
<td>3 (75%)</td>
<td>3 (75%)</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>V2 - ulɔ</td>
<td>3 (75%)</td>
<td>1 (25%)</td>
<td>4 (50%)</td>
</tr>
<tr>
<td>V3 - ulaf</td>
<td>2 (50%)</td>
<td>0 (0%)</td>
<td>2 (25%)</td>
</tr>
</tbody>
</table>
Figure 2: The distribution of phonological variants [ular] in Northern Perak.

The map in Figure 2 above shows the existence of phonological distributions of [ular]. Values 1, 2 and 3 on the legend refer to the number of pronounced variants based on the district respectively. The peach colour refers to the area which has only one variant that is [ula], [ulo] or [ulo[i]. The blue colour refers to the area which has two pronounced variants, [ula] and [ulo]. Whereas the purple colour refers to the areas which have three pronounced variants. All of these variants are used in different areas. The variant [ulo[i] is very dominant and is productively used in Larut Matang & Selama (LMS). [ulo[i] is distributed from the western areas and is strongly influenced by the northern dialect, namely Kedah. Interestingly, the variant [ula] has a wider distribution. [ula] is used both in LMS and Hulu Perak (HP) occurring in Kerunai, Lenggong and Pengkalan Hulu in the HP district and Trong, Bukit Gantang in the LMS district. These areas are lowlands. [ula] is clearly influenced by Pattani dialect from the northern and eastern neighbouring borders. Finally [ulo] was found mostly at Larut Matang and Selama namely at Kampung Pauh, Kampung Ayer Kuning, Kampung Sempeneh Sebarang, Kampung Sempeneh Cempaka and Kampung Kubu. From the map drawn above, we can see that the most complex area is in LMS district especially in Bukit Gantang and Trong sub-district (denoted by ★). All variants [ulo], [ula] and [ulo[i] are used in this area.
The variants for [u₁ə], [u₂a] and [u₂u] rely on topographical features such as highlands, neighbouring borders, river basins and river. These interesting facts can be explained better with GIS. The combination of different variables such as speakers, boundaries, topographies are major factors that led to these occurrences. For instance, Trong and Bukit Gantang share a major road that connects these two places. Geographically, both areas are bordered to the west by Kedah. Trong is adjacent to the sea whereas Bukit Gantang has an access road to Kuala Kangsar, a lowland area and a centre for royal administration. Even though there is a highland near Bukit Gantang that separates Bukit Gantang from Kuala Kangsar, there is still an access road to Kuala Kangsar. Hence the fluidity of people is substantial. Consequently, activities that allow people to interact contribute a great impact to the existence of the variants. As for Pengkalan Hulu, borders that form the west, north and east have once again given an impact to the distribution of the variants.

5. Conclusion

A multidisciplinary approach such as geo-linguistics has recently become more popular. GIS which was once restricted to geography has been widely applied to other disciplines as well. It has become useful for information system analysis especially for data mining. In addition, GIS is now a friendly tool for researchers from other disciplines to use. Research on the lexicals, [mereka] and [ular] in northern Perak clearly proves that GIS can help improve the presentation of dialects distribution more systematically.

References