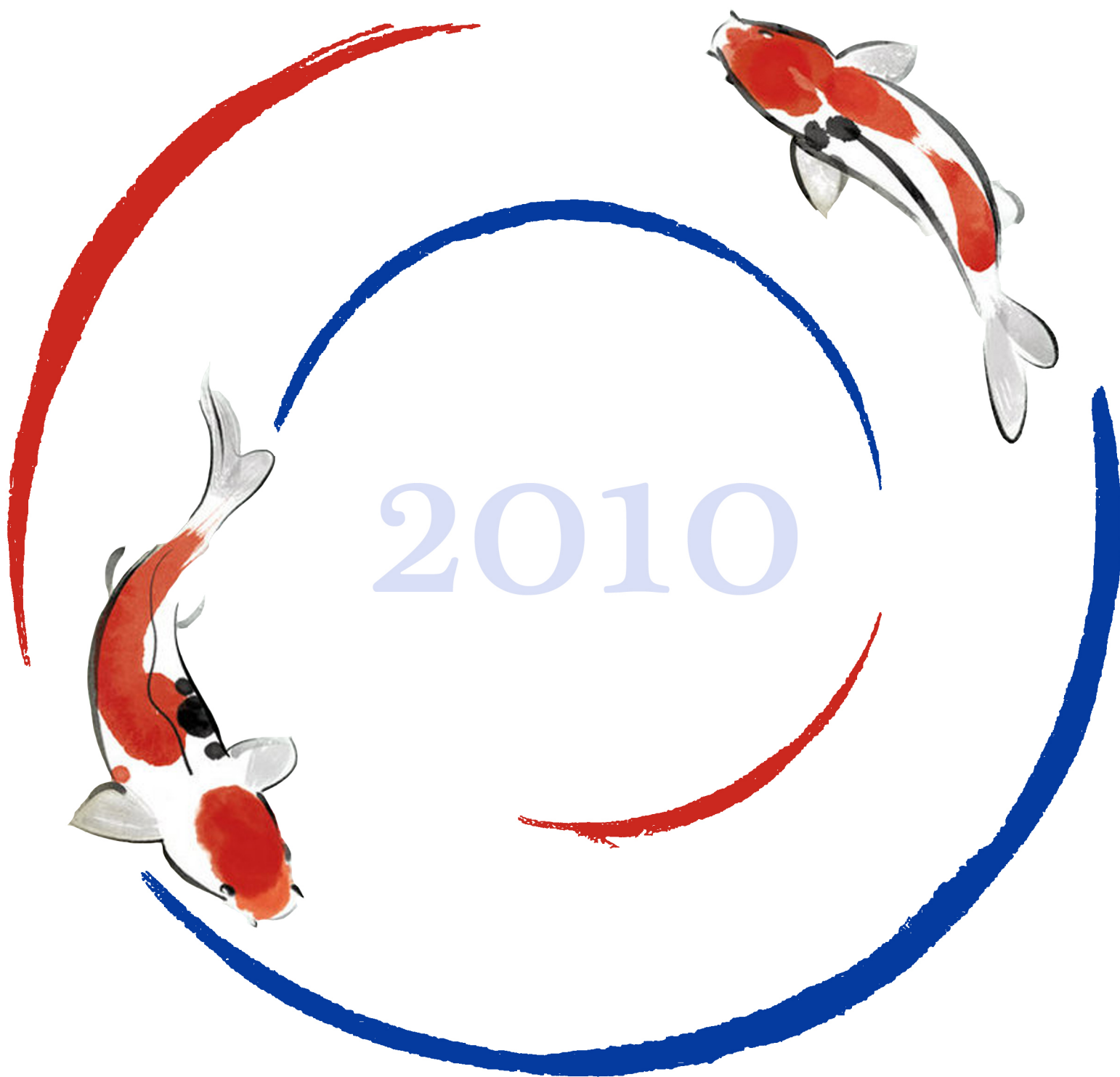


acah2010

The Asian Conference on Arts and Humanities

acss2010

The Asian Conference on the Social Sciences



June 2010, Osaka, Japan
Official Conference Proceedings

iafor

COVER PAGE

Topic: A Geographical Look at the Word Geography of Thailand

(reference abstract no.: 0133)

Presenter: Sirivilai Teerarojanarat, Ph.D.

Department of Geography, Faculty of Arts, Chulalongkorn University,
Phayathai Road, Bangkok 10330, Thailand

Sirivilai.t@chula.ac.th

Asst. Prof. M.R. Kalaya Tingsabadh, Ph.D.

Department of Linguistics, Faculty of Arts, Chulalongkorn University,
Phayathai Road, Bangkok 10330, Thailand

kalaya.t@chula.ac.th

Affiliation: Chulalongkorn University, Thailand

INTRODUCTION

Many studies of Thai dialectology have been carried out over the past three decades. In previous studies, however, maps were drawn manually on papers, the locations of data collection as well as the drawing of isoglosses and dialect boundaries were roughly marked. The maps created in that way consequently have small degree of positional accuracy and lack reliability.

Thanks to technological inventions from the West, the advancement of geographical tools and technologies in the last few decades, especially through Geographic Information System (GIS) and Remote Sensing (RS), have been developed to enhance the ways to observe, collect, measure, and analyze spatial data as well as to display and produce maps more efficiently. The derived map information is now widespread in a digital form and is sharable among the digital world.

In 2002 the Word Geography of Thailand project was launched under the sponsorship of the Ministry of Culture with its aim to collect, record, and set up a database of Thai dialect vocabularies through the study of lexical items. The project pioneered in applying GIS to develop the geographical database of lexical items in Thai dialects. A part of the work results was displayed as a set of 170 lexical variation maps. These dialect maps shaded polygons on the basis of subdistrict boundaries to illustrate the dialect distribution. These maps also showed the interesting distribution of lexical items all over the country i.e. the boundaries between the four main dialects as well as the existence of subdialect and local dialect areas.

Although the project has achieved great benefit from the geographical database, shading maps on the basis of polygons is, however, aggregated to political boundaries and unavoidably uniform. It is, to some degree, unrealistic as it cannot truly reflect the distribution in reality. The extended work is demonstrated in this paper. Its key aim is to present an alternative look by which GIS and geographic features are incorporated to improve the pattern of dialect distribution to be more realistic. Geographic features - population settlement, topography, and transportation - are included for investigation and analysis.

STUDY AREA

Location of the study area covers the whole Thailand. Geographically this country, lying in the heart of the Southeast Asian mainland, covers an area of about 514,000 sq. km. or roughly the size of France. Thailand borders Myanmar and Laos to the north, Cambodia and the Gulf of Thailand to the east, Myanmar and the Indian Ocean to the west, and Malaysia to the south. In this country, more than 60 ethnolinguistic groups with their diversity of languages were reported (Suwilai and others, 2004). Despite the language diversity, Thai, classified by linguists as belonging to Chinese-Thai branch of the Sino-Tibetan family, is the official national language spoken in every part of the country. This language, differentiated by the areas of the country where dialects are spoken, can be divided into four main dialects comprising Northern

Thai, Northeastern Thai, Central Thai, and Southern Thai. Spatial variation of these four main dialects is the main focus of this study.

BACKGROUND TO PRIMARY DATA SOURCE: THE WORD GEOGRAPHY MAPS OF THAILAND

The primary data source of this study was from the geographical database of the Word Geography of Thailand project, namely the Word Geography maps of Thailand. Data covers the whole of Thailand at sub-district level, so-called “tambon” in Thai, except the Bangkok Metropolis. A questionnaire with 170 questions - each representing a semantic unit - was constructed. The questionnaire was sent in 2003 to all of the 7,226 tambons via the Ministry of Culture network in the first round and by post in the second round. The data used in this study came from 6,379 tambons or 88% of the study locations. The respondents were a director or a teacher of a tambon school. Data collection was completed within 1 year (Sirivilai and Kalaya, 2008).

Linking lexical data and administrative boundary maps was designated and performed under a GIS environment using a relational database management system (RDBMS) as shown in Figure 1. From the figure, each word map contains a set of lexical items. For ease of use, visual interpretation, and model simplification, it is assumed that each tambon has a uniformed lexical usage. Thus, a lexical item chosen for each semantic unit is used to be a representative of that whole tambon (in the form of polygon). Words that contain minimal phonological discrepancies e.g. /maak^T mii^T/ and /mak^T mii^T/¹ are analyzed as a single lexical item. A geographical database, as a result, contains 170 lexical variation maps - one map per one semantic unit. As the four main dialects - Northern Thai, Northeastern Thai, Central Thai, and Southern Thai - were the main focus, the lexical items that accumulate in one region, displaying membership of the same variety, were assigned in different shades of the same colour i.e. magenta shades for Northern Thai, blue shades for Northeastern Thai, yellow shades for Central Thai, and green shades for Southern Thai. An example of word maps is shown in Figure 2. The whole dataset of 170 lexical variation maps can be currently visited through the website Word Geography Maps of Thailand (<http://www.arts.chula.ac.th/~ling/geoling/WGT/>).

¹ In this study tones are marked as T and excluded in the lexical analysis since tonal variation in Thai is highly complex and has to be studied separately (Tingsabadh, 2001)

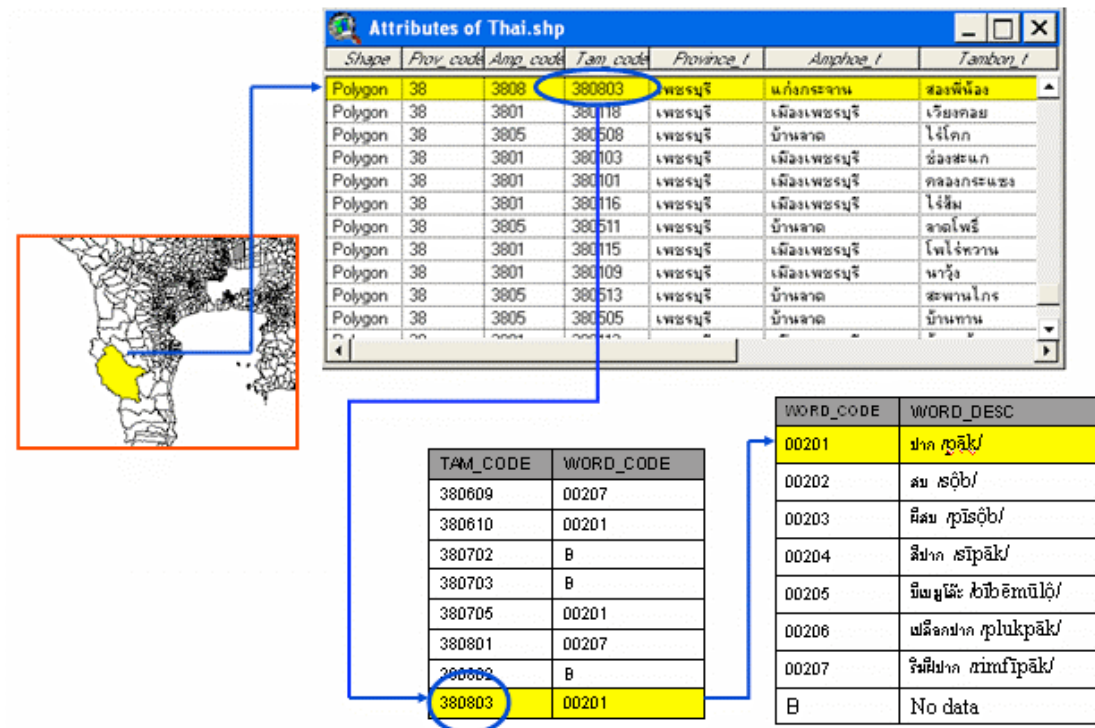
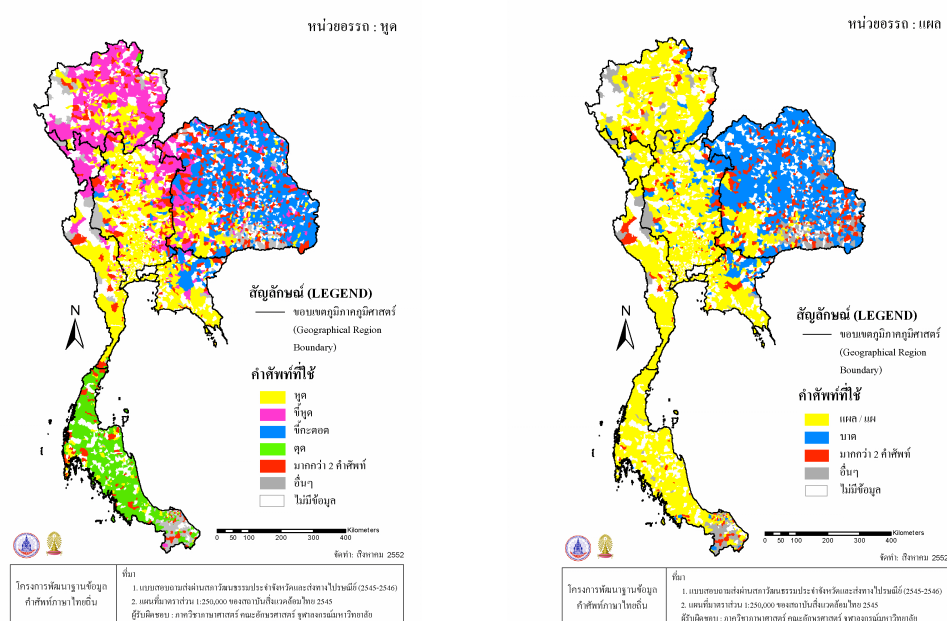


Figure 1: An example of linkage product – the administrative map and the lexical data. Here is a record of tambon “Song Phi Nong” being highlighted in reference to the lexical item /paak^T/ “mouth”.



(a) Semantic unit “wart”

(b) Semantic unit “wound”

Figure 2: Word maps of the semantic units (a) “wart” and (b) “wound”.

GEOGRAPHIC FEATURES AND THEIR ROLES FOR THE ANALYSIS

Geographic features play key roles in this analysis. Three types of demographic and geographic features were included in this study. All are available in digital format. They are population settlement, topography, and river transportation.

According to the fact that language is closely bound up with humans, population settlement, then, acts to locate and shape the pattern and distribution of where dialects are spoken. Population settlement in this study was derived from the 2005 Landscan population distribution model. The model was produced under the Oak Ridge National Laboratory (ORNL) project, by the United States Department of Energy's (USDOE) Oak Ridge National Laboratory (ORNL). The model, analysed by the technologies of GIS and RS, contains the available census counts (usually at sub-province level) for each country and four main physical datasets, namely land cover, road, slope, and night time light obtained from the satellite imageries in order to be used as key indicators to produce a population distribution map. Nowadays Landscan is considered as having the most up-to-date and finest global population data with the spatial resolution of approximate 1 km. Landscan data covers the entire global area and can be applied to various applications that require geographically referenced population bases. For more information about the Landscan data, visit website <http://www.ornl.gov/sci/landscan/>.

The remaining types of geographic features – topography and river transportation, acted to influence and shape the earliest patterns of settlement and migration and to initiate the dialect differences in each region. The effect of topography e.g. mountain ranges and mountainous areas has long been investigated and assessed in many dialect studies to account for its role on the separation between major dialects and/or subdialects e.g. the work of (Garr, 1985; Luo et al., 2009). Topography was available in the study in the raster form of digital elevation model (DEM). It was obtained free of charge from the USGS (the U.S. Geological Survey), namely the SRTM (the NASA Shuttle Radar Topographic Mission) digital elevation data. The SRTM Data, originally produced by NASA, covers all countries of the world at the approximate spatial resolution of 1 km. For more information about the SRTM data, visit the CGIAR-CSI website <http://srtm.csi.cgiar.org/>. Apart from the topographic data, transportation through the study of waterways (e.g. rivers and streams) plays a role in locating the origin and cluster and spread of population settlements in the early days (Luo et al., 2000; 2006; 2009). River data is available in the vector format, obtained from the Department of Transportation, Thailand.

METHODOLOGY

Analysis of the study was divided into two parts. The first part involved the way to shape the dialect pattern while the second part mainly examined the effect of geographic features on the separation of major Thai dialects in each region.

In the first part, all word maps were first re-structured and geo-coded in GIS-based raster grid cells (of about 1 km x 1 km) to be corresponded with that of a topographic map. Next step is to define the extent of population settlement. Using the population data from Landsat, population density was then chosen to demarcate the settlement area. Setting criteria was carried out with the simple requirement that the number of population density has to be accurate enough to map the extent of most urban and rural settlements in the study area. It should be noted that up to now definition, terminology and criteria for classification of populations in urban/rural settlement areas are not in common. Setting the criteria varies extensively in different countries e.g. United Kingdom using the population of $\geq 10,000$ persons/sq km to locate the settlement area (UNSTAT, 2010). In case of Thailand, municipal areas are used as an index (UNSTAT, 2010). The agglomeration of $\geq 10,000$ inhabitants with the population density of $\geq 3,000$ inhabitants/sq km is used to define a town while the agglomeration of $\geq 50,000$ with the population density of $\geq 3,000$ inhabitants/sq km is used to define a city. Among dissimilarities in terminology, this study followed the definition of Organisation for Economic Cooperation and Development (OECD, 1994) in that a “rural area” is determined roughly by having a population density < 150 persons/sq km. By using this number as a reference, this study further visually observes and examines the pattern of settlement, and then adjusts the level of population density to the appropriate number. Figure 3 shows a series of settlement areas as a result of the adjustment of different levels of population density. Finally, the population density of ≥ 50 persons/sq km was then chosen to demarcate the settlement area. The next step is then to use the previously delineated settlement area to extract the dialect patterns from the word maps. The shape and location of where dialects are spoken in the study area are the ones plotted last.

In the second part, the effect of geographic features – topography and waterways - were superimposed on the new word maps previously produced. Examination was carried out on the basis of visual observation. Results and discussion are given in the next section.

RESULTS AND DISCUSSION

Results of the analysis were 170 word maps, each representing one semantic unit and shaped with population settlement. Figure 4 demonstrates a part of a word map ‘before’ and ‘after’ shaping with population densities in comparison. From the figure it is obvious that shapes and patterns of the ‘before’ and ‘after’ word maps are dramatically different. The word map in Figure 4(a) is shaded uniformly. From an overall look, the map helps depict the lexical variation all over the entire subdistrict area. The resultant map, after being shaped with population density of ≥ 50 persons/Sq km, in Figure 4(b) shows some improvement over Figure 4(a) in displaying the true locations of where people use dialects. The spatial variation and concentration of lexical items was also revealed. Furthermore, draping the word map in Figure 4(b) with topographic features as shown in Figure 4(c), can help linguists to better understand the relationship between dialect phenomena and its physical environment in the study area. Figure 5 also shows the overall picture of a resultant word map, here with the exemplified semantic unit “to belch”, with the magnified geographic regions of the country; the North, the Northeast, the Central, and the

South. The different look of the word map suggests another way, by incorporation of geographic features, to improve the presentation of lexical variation in the real world. In addition, the map results can help linguists have more confidence in interpreting the results, drawing isoglosses, and/or language boundaries.

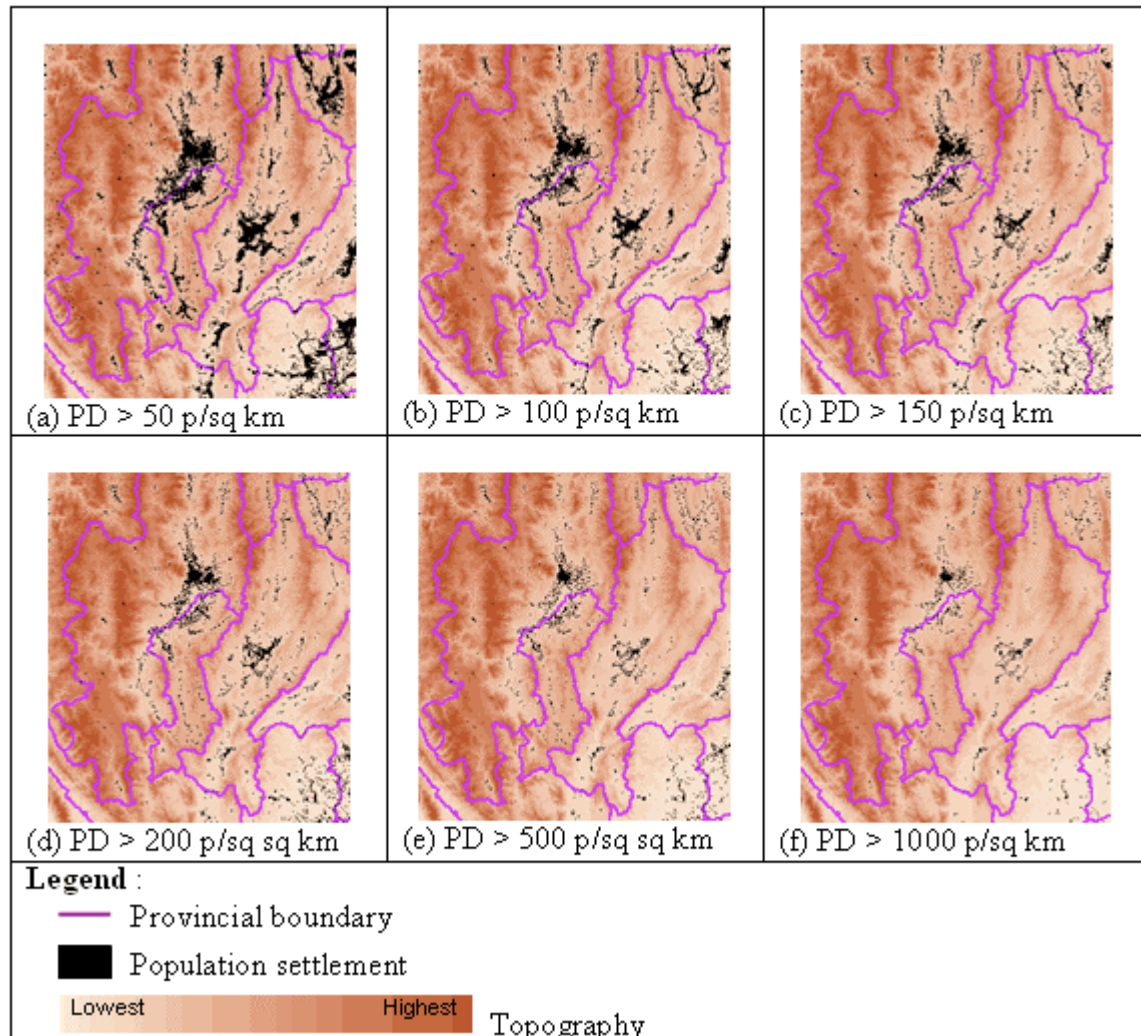


Figure 3: Population settlement with different levels of population density (PD) from (a) population density greater than 50 persons/sq km to (f) population density greater than 1,000 persons/sq km, here draping with topographic features.

Despite the beneficial use of this technique, what one should be aware of is that different criteria for demarcating population settlement can result in different shapes and patterns of word map presentation. In Figure 4(b), for instance, once the specified population settlement was shaped, it completely omitted the grey color which can be obviously seen in the original map in Figure 4(a). This means that the grey area is where population density is less than 50 persons/sq km. That is, if the study wants to cover this grey area, the level of population density needs to be adjusted.

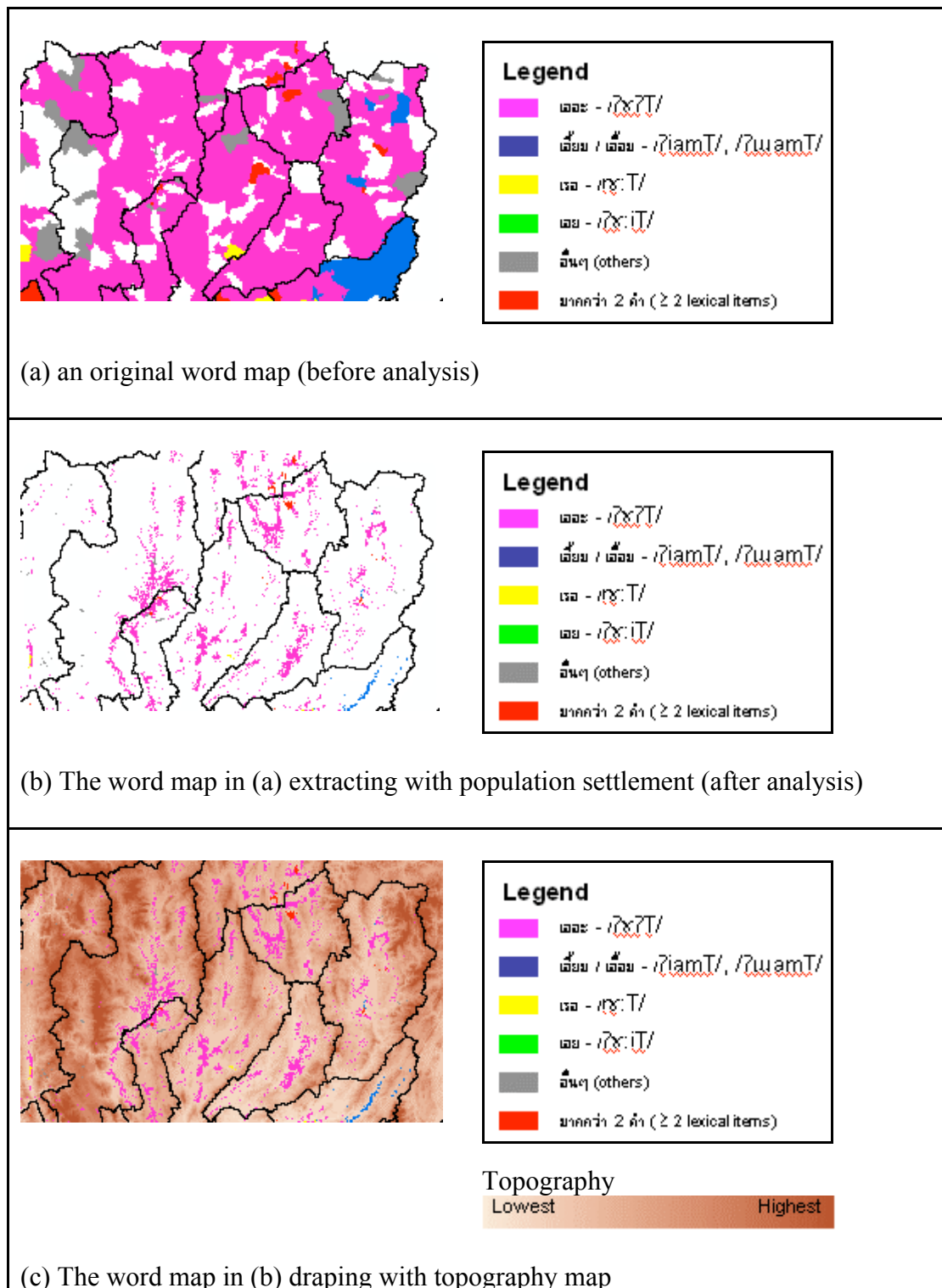


Figure 4: A part of word map of a semantic unit “to belch” showing the (a) ‘before’ and (b) ‘after’ shaping with population density of ≥ 50 persons/sq km, and (c) that of (a) draping with topography.

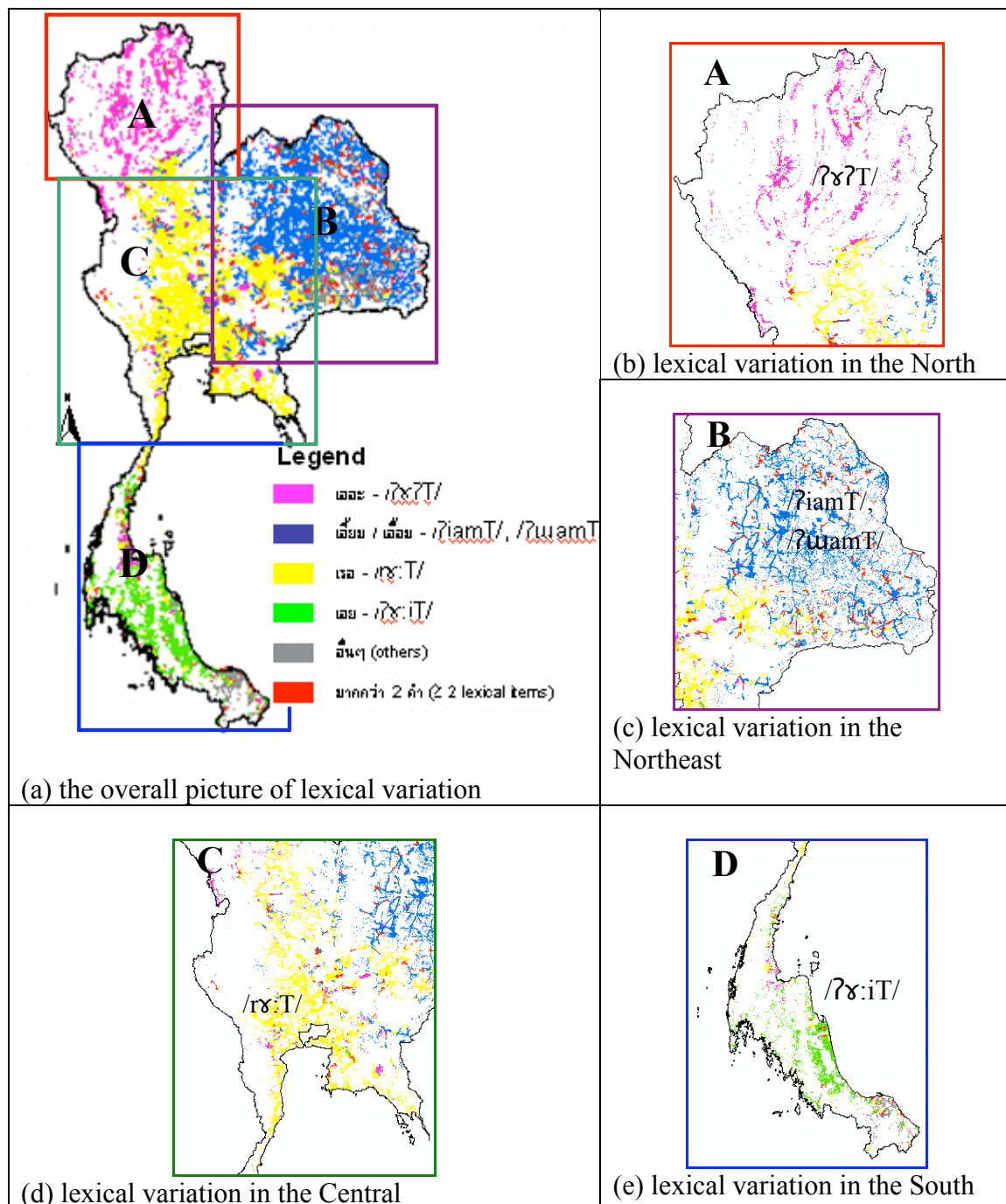


Figure 5: A word map of the semantic unit “to belch” with the magnified area in (a) the North, (b) the Northeast, (c) the Central and (d) the South.

Next, the effect of topography and waterways was investigated through visual observation. In this study area, three noticeable points can be concluded. Firstly, topography plays a key role for the separation between Northern-Central Thai and Northeastern-Central Thai, but not in the Southern-Central Thai. In spite of their

dissimilarities in lexical distribution² as example maps in Figure 6, overall the Northern-Central Thai and the Northeastern-Central Thai boundaries are quite clear and stable. These phenomena can be explained through the investigation of topography. Figure 7 shows the characteristics of the topography of Thailand which can be mainly divided into four natural regions; (1) the North, (2) the Central Plain, (3) the Northeast, and (4) the South. The North is a region dominated by mountains, natural forests, ridges, narrow and fertile valleys. The Central Plain, so-called Chao Phraya River Basin, is a flourishing fertile valley. This region is the richest rice-producing area of the country, irrigated by the Chao Phraya River. Bangkok, the capital of Thailand is located in this region. The Northeast, so-called “Isan” in Thai, is characterized by the Korat Plateau, a rolling surface and undulating hills, watered by the Shee and Mun River. The South is characterized by a long narrow peninsula comprising a hilly to mountainous region. Due to the topographic differences the mountainous region of the North and the Korat plateau of the Northeast can be visibly distinguished from the flat area of the central region as shown in the Figure as a blue dash line. This finding thus confirms the effect of topography on the separation of three major dialects - Central Thai, Northern Thai and Northeastern Thai. In case of Southern-Central Thai in which the topographic characteristics are lacking, other factors, especially historical and cultural contexts e.g. the boundaries between old administrative areas and cultural differences, should be further investigated.

Secondly, waterways, especially through rivers and water basins, present favorable physical environments for settlement which leads to the cluster of one dialect spreading over a large area, especially one clear cluster in the central region and another cluster in the northeastern region in Figure 6 as well as the cluster of subdialects along the waterways in the northern region and the cluster of Southern Thai around the lake, so-called “Songkhla Lake” basin³, in the southern region as shown in Figure 8.

Finally, through the overlay of a topographic map the separation between subdialects/local dialects can be seen and explained through the effect of topography and waterways in the northern region as shown in Figure 8 (left). However, the effect on the separation between main dialects and local dialects in the southern region (see Figure 8 (right)) cannot explain to the overlay of a topographic map. Suggestions to include other factors such as main roads and railways to the analysis is required for future investigation. In addition, it should be noted that, overall, the areas where subdialects and local dialects are located are not clearly seen in this study as these areas have population densities of < 50 inhabitants/sq km. It is thus recommended that

² One significant finding of the previous project, the Word geography of Thailand, is that each semantic unit has its own independent pattern of lexical variation. Some maps show the clustered pattern – the homogeneity of lexical items over a large area, whereas some show the dispersion pattern - a mixture of lexical items covering an entire area. Also some word maps appear as four clear clusters, some appear as three or two clusters, while at least one map appears as a single cluster due to lack of variation.

³ Songkhla Lake is the largest natural lake of Thailand, covering an area of about 1,000 sq km.

for a future study in which subdialects and local dialects are the main focus, adjusting the level of population density is needed.

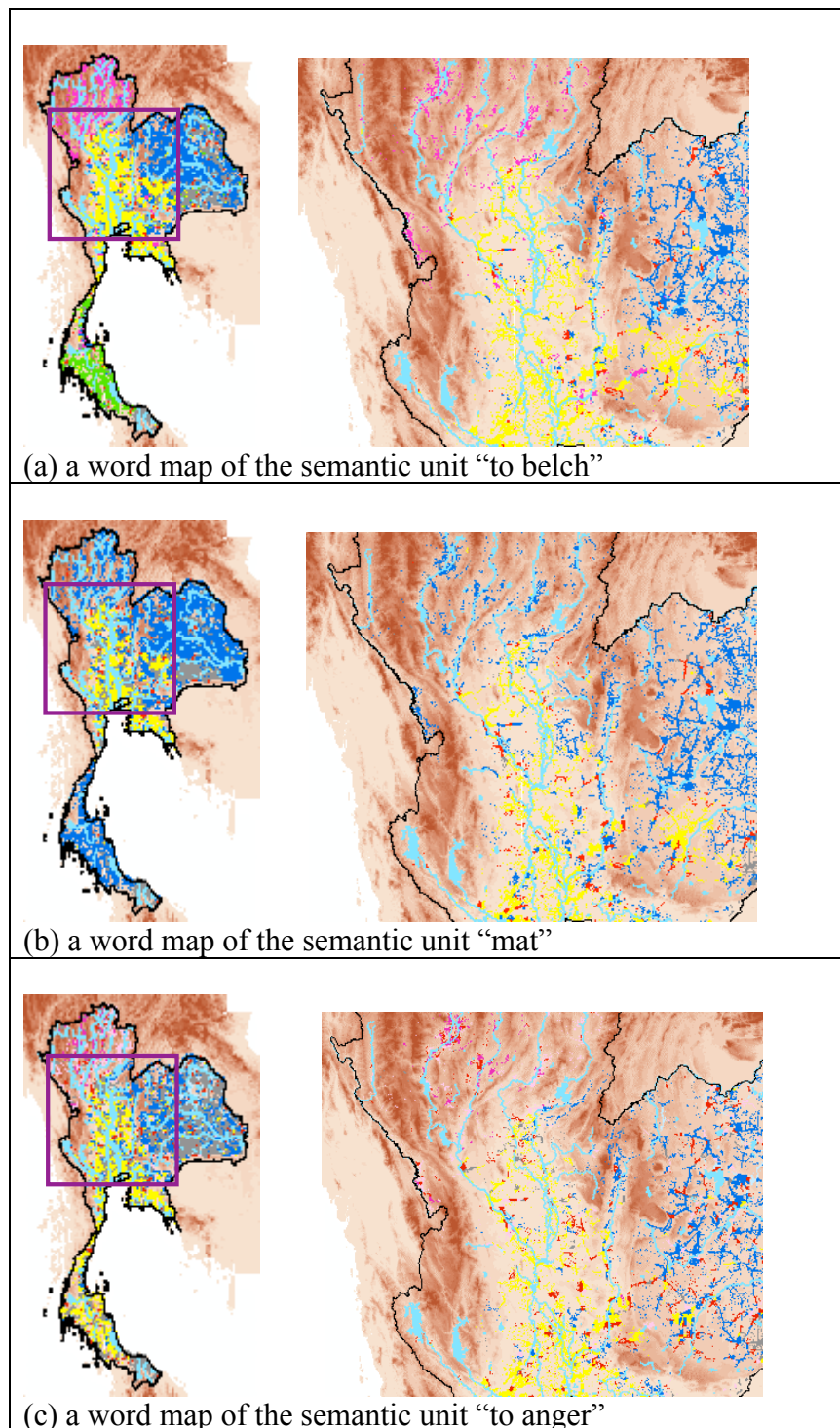


Figure 6: Word maps of the semantic unit (a) “to belch”, (b) “mat” and (c) “to anger” in comparison.

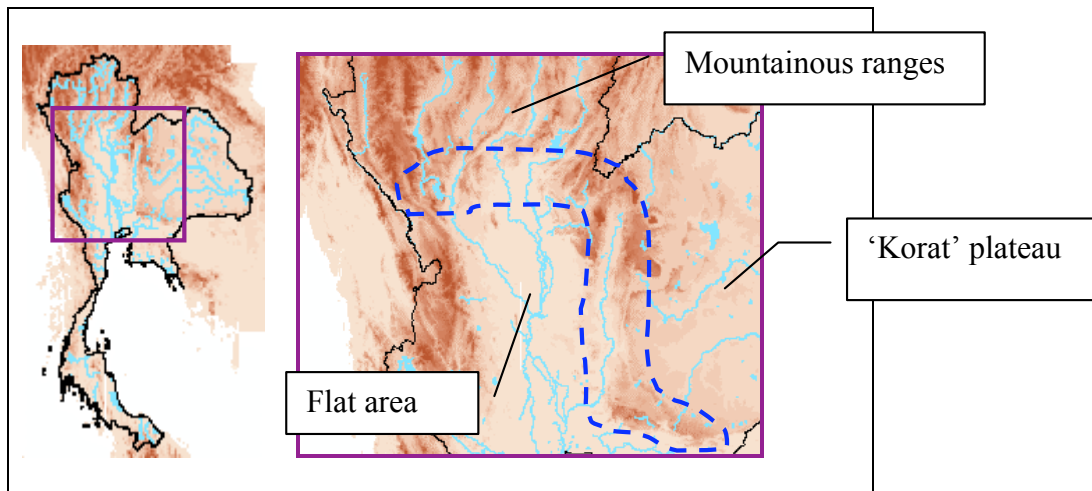


Figure 7: The topographic characteristics of Thailand with the magnified area of central region.

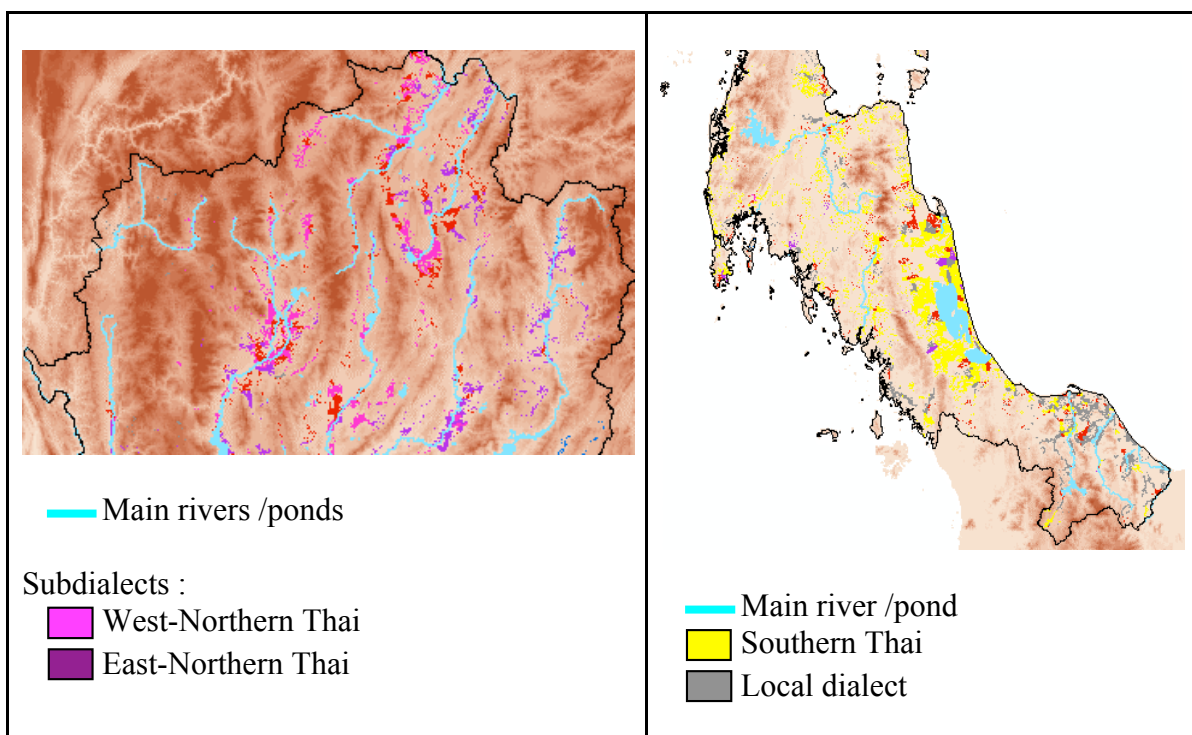


Figure 8: A word map of the semantic unit “to anger”

CONCLUSION AND FUTURE RESEARCH DIRECTION

This study shows how GIS analysis and incorporation of the demographic and geographic features – population settlement, topography, and river transportation – can help improve the pattern of dialect distribution to be more realistic. Results show that shaping dialect maps with population settlement helps explore the dialect patterns and their locations more accurately. Results also reveal the obvious effect of

topography on the separation of major Thai dialects, especially between Northern-Central Thai, between Northeastern-Central Thai, as well as between subdialects in some regions. It also shows the clustering pattern of dialect on the effect of transportation through waterways. The results thus suggest that the analysis gives a geographical perspective which helps detect the spatial pattern of dialect distribution. It also bridges the fundamental understanding between the fields of linguistics and geography in order to explain the impact of the physical environments on human settlement and activities. Further analysis is required, especially the reclassification of population settlements which appear mostly in the border regions of the country inhabited by many ethnic minority groups. Another research direction is to include roads and railways to investigate their impact on dialect change we see today.

REFERENCES

- Garr, W. R. 1985. *Dialect Geography of Syria-Palestine: 1000-586 B.C.E.* Philadelphia: University of Pennsylvania Press, 1985.
- Luo, W., Hartmann, J. F., Li, J., and Sysamouth, V. 2000. GIS Mapping and Analysis of Tai Linguistic and Settlement Patterns in South China. *Geographic Information Sciences*. 6(2): 129-136.
- Luo, W., Hartmann, J. F., Wang, F. 2009. Terrain characteristics and Tai toponyms: a GIS analysis of *Muang*, *Chiang* and *Viang*. *GeoJournal*. DOI 10.1007/s10708-009-9291-8. (<http://www.springerlink.com/index/5842885033597132.pdf>) Accessed 1 February 2010.
- OECD (1994) Creating Rural Indicators for Shaping Territorial Policies, OECD Publications, Paris, France.
- United Nations Statistics Division (UNSTAT). (<http://unstats.un.org/unsd/demographic/sconcerns/densurb/densurbmethods.htm>) Accessed 1 February 2010.
- Wang, F., Hartmann, J., Luo, W., and Huang, P. 2006. GIS-Based Spatial Analysis of Tai Place Names in Southern China: An Exploratory Study of Methodology. *Geographic Information Sciences*. 12(1): 1-9.
- Sirivilai Teerarojanarat and Kalaya Tingsabadh. 2008. Development of Geographical Database of Lexical items in Thai Dialects. *Journal of Language and Linguistics*. Department of Linguistics, Faculty of Liberal Arts, Thammasat University 26(2): 1–20. (in Thai)
- Suwilai Premsrirat and Others. 2004. *Ethnolinguistics Maps of Thailand*. Bangkok: Office of the National Culture Commission in Ministry of Culture. (in Thai)
- Tingsabadh, K. (2001) “Thai Tone Geography”. In *Essays in Tai Linguistics*, pp 205-228, edited by Kalaya Tingsabadh and Arthur S. Abramson. Bangkok: Chulalongkorn University Press.

The Landscan website: <http://www.ornl.gov/sci/landscan/>. Accessed 1 February 2009.

The CIAT SRTM website : <http://srtm.csi.cgiar.org/>. Accessed 1 February 2009.

The Word Maps Geography of Thailand website:
(<http://www.arts.chula.ac.th/~ling/geoling/WGT/>). Accessed 1 February 2009.

ACKNOWLEDGEMENTS

The authors wish to thank the following agencies who provided data source for the analysis; dialect data from the Word Geography of Thailand, administrative boundaries map of Thailand from the Thailand Environment Institute (TEI), river map from Department of Transportation, Thailand, Landscan data from the Oak Ridge National Laboratory (ORNL) project, and Digital Elevation Model (DEM) from the U.S. Geological Survey (USGS). We are also very grateful to Chulalongkorn University and the Faculty of Arts, Chulalongkorn University for the financial support that made this study possible. Without their support this study would not have been completed.