

## **Word Geography of Thailand**

M.R. Kalaya Tingsabadh\*, Phinnarat Akharawatthanakun, Uraiporn Tantinimitrkul,  
Sirivilai Teerarojanarat, Ornusa Chinkrachangkit

Faculty of Arts, Chulalongkorn University,  
Phyathai Road, Bangkok, Thailand  
kalaya.t@chula.ac.th

### **Introduction**

Thai is spoken in every part of Thailand. Lexical variation and tonal variation are the main factors that differentiate the four main dialects of this language: Northern Thai, Northeastern Thai, Central Thai, and Southern Thai. The names of the dialects indicate the areas of the country where these dialects are spoken.

Many studies of both lexical and tonal variation of Thai have been carried out over the past three decades. As far as lexical studies are concerned, this Word Geography of Thailand project differs from all of the previous studies in three ways. Firstly, it covers the whole of Thailand. Most of the previous projects covered an area of just one or two provinces (e.g. Duangjai 1985, Vichintana 1986) while the whole region and the whole country were investigated in just a few studies (Wantanee 1983, Jaroon and Wantanee 1983, Chalida 1990). Secondly, while the previous studies collected data at the level of amphoe or district, this study collects data at the level of tambon, which consists of villages and is a lower level administrative unit than amphoe. Thirdly, map production in this study employs Geographical Information System (GIS). In previous studies maps were drawn manually on paper, the locations of data collection as well as the drawing of isoglosses were roughly marked. The maps created in that way consequently have small degree of precision and lack reliability. In this study map quality is very high allowing better interpretation of results.

### **Methodology and Analysis**

This project covers the whole of Thailand, except the Bangkok Metropolis, which comprises 7,226 tambon. A questionnaire with 170 questions - each representing a semantic unit - is constructed. The selection of the semantic units is based on the results of several previous studies. The ones included in this project display the isoglosses that have high potential in differentiating dialects and/or subdialects.

The questionnaire was sent in 2003 to all of the 7,226 tambon 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 tambon 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.

Geographical Information System (GIS) is used in this study to overcome or, at least, lessen the drawbacks that occur in the previous studies. GIS is a computer system designed for managing geographic data (Heywood et al., 2002; Longley et al., 2005). Distinct from the MIS (Management Information System) such as the tabular data of Microsoft Access and

Oracle, GIS provides a database linking the geographic features to their attributes and the capabilities of data collection, storage, retrieval, display, and analysis. GIS was applied in this project for two main purposes. The first was to develop the geographical database for storing lexical items. The second one was to create lexical variation maps for the 170 semantic units.

To achieve these two purposes, two main data sources were used as the following.

1. Lexical tabular data of 170 semantic units. The data is stored in a relational database of Microsoft Access (the lexical MIS database).

2. A Thailand's administrative boundary map at the scale of 1:250,000. The map, used as the base map of this project, was obtained from Thailand Environment Institute (TEI), Thailand. The map is a shapefile (vector data model). Features are stored as polygons having the detail of district levels.

Linking lexical data and administrative boundary map (shapefile) was done within a GIS environment using the GIS software, ArcView version 3.2. Figure 1 shows the relationship between the lexical MIS database and the lexical GIS database developed. Lexical data of 170 semantic units stored in the lexical MIS database was then converted and transferred to the GIS software through Microsoft Excel. For each semantic unit, this shapefile was then permanently linked to lexical items to create a new lexical variation map.

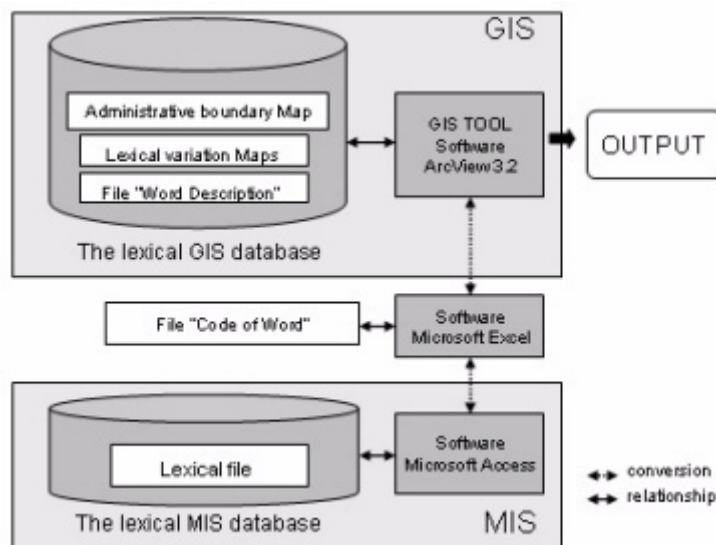


Figure 1: Database of the project.

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). In order to establish the relationship, field 'TAM\_CODE', being used to store the code of tambon, is then used as a key index to link the map with lexical tabular data (see an example in Figure 2).

At this stage a preliminary lexical analysis is carried out. Words that contain minimal phonological discrepancies e.g. /maak<sup>T</sup> mii<sup>T</sup>/ and /mak<sup>T</sup> mii<sup>T</sup>/<sup>1</sup> are analyzed as a single lexical item.

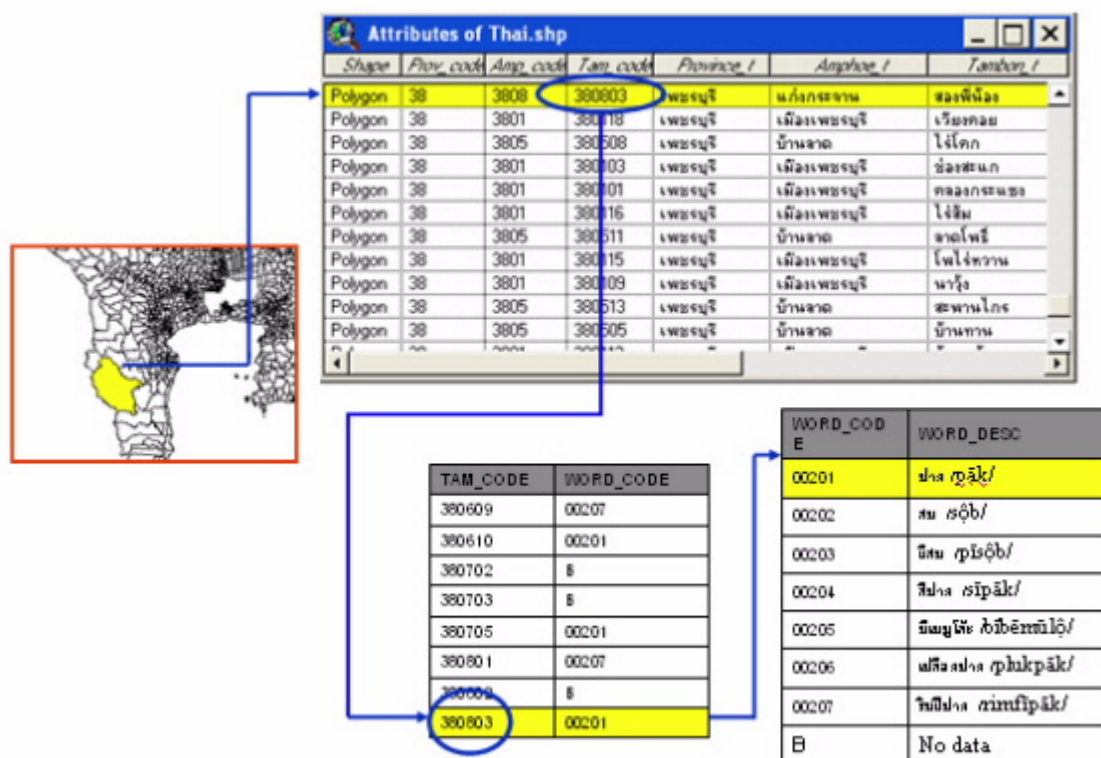
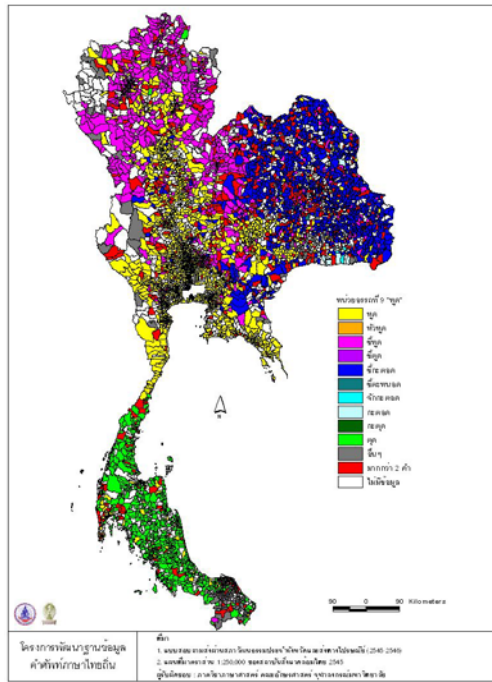


Figure 2: An example of linkage product – the administrative map and the lexical data. Here one record of a tambon in Song Phi Nong district was retrieved. It was coded ‘00201’ referring to the lexical item /paak<sup>T</sup>/ “mouth”.

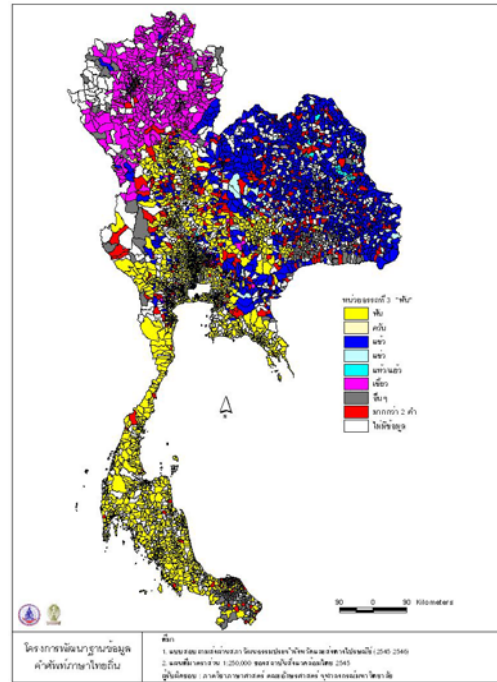
A geographical database, as a result, contains 170 lexical variation maps - one map per one semantic unit. Each resultant map contains a set of lexical items stored as codes in a ‘WORD\_CODE’ file (see Figure 2). Each also temporarily links to “Lexical description” file to describe lexical codes. These maps can be displayed spatially as four examples shown in Figure 3.

Analysis is carried out in two steps. In the first step one colour is assigned to one lexical item. Each map produced at this stage freely displays lexical variation in the whole country. In the second step the four main dialects - Northern Thai, Northeastern Thai, Central Thai, and Southern Thai - are the main focus. The lexical items that accumulate in one region, displaying membership of the same variety, are assigned in (different shades of the same color) i.e. magenta shades for Northern Thai, blue shades for Northeastern Thai, yellow shades for Central Thai, and green shades for Southern Thai. Subdialects are displayed when distinct areas appear in different shades of a single colour.

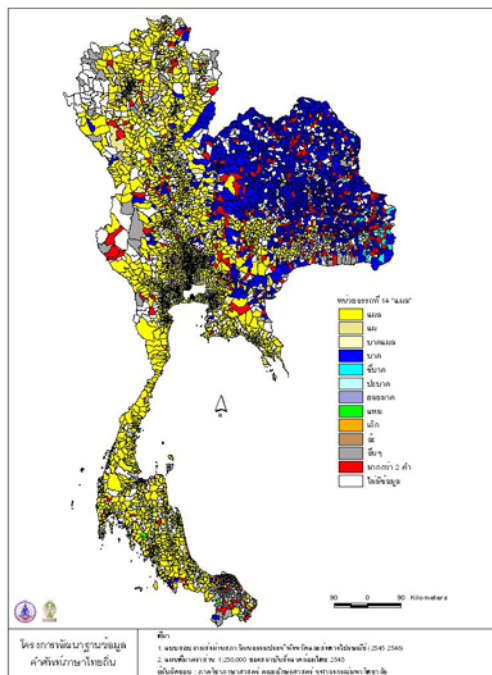
<sup>1</sup> In this study tones are marked as T and excluded in the lexical analysis since tonal variation in Thai is highly complex and has to be studied separately (Tingsabadh, 2001)



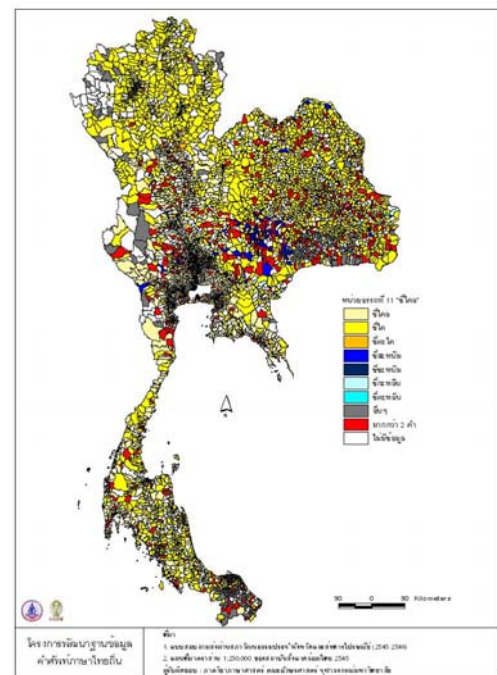
(a) Semantic unit “wart”



(b) Semantic unit “teeth”



(c) Semantic unit “wound”



(d) Semantic unit “sweat stain”

Figure 3: A lexical variation map of the semantic units (a) “wart”, (b) “teeth”, (c) “wound”, and (d) “sweat stain”

## Result and discussion

By visually observing the 170 lexical variation maps, two noticeable points can be concluded. Firstly, 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 shows the dispersion pattern - a mixture of lexical items covering the entire area. A mixture of these two main patterns is also found frequently. Such dissimilarities can be explained as the uniqueness of every single map. This finding is in line with the saying that “each word has its own history”. Figure 3 shows different patterns of lexical variation of four semantic units. As clearly seen, while the semantic unit “wart” appears as four clear clusters, the semantic unit “teeth” appears as three clusters, and the semantic unit “wound” appears as two clusters. The semantic unit “sweat stain”, however, appears as a single cluster due to lack of variation.

Despite their dissimilarities in lexical distribution, overall the Northern-Central Thai and the Northeastern-Central Thai boundaries are quite clear and stable (see Figure 4 (a) and (b) respectively), while the Central-Southern Thai one varies from one semantic unit to another (see Figure 5). Moreover, some of the maps show the area of subdialects and local dialects (see Figure 6 (a)). At least one map shows how lexical change comes with new style of living (see Figure 6 (b)).

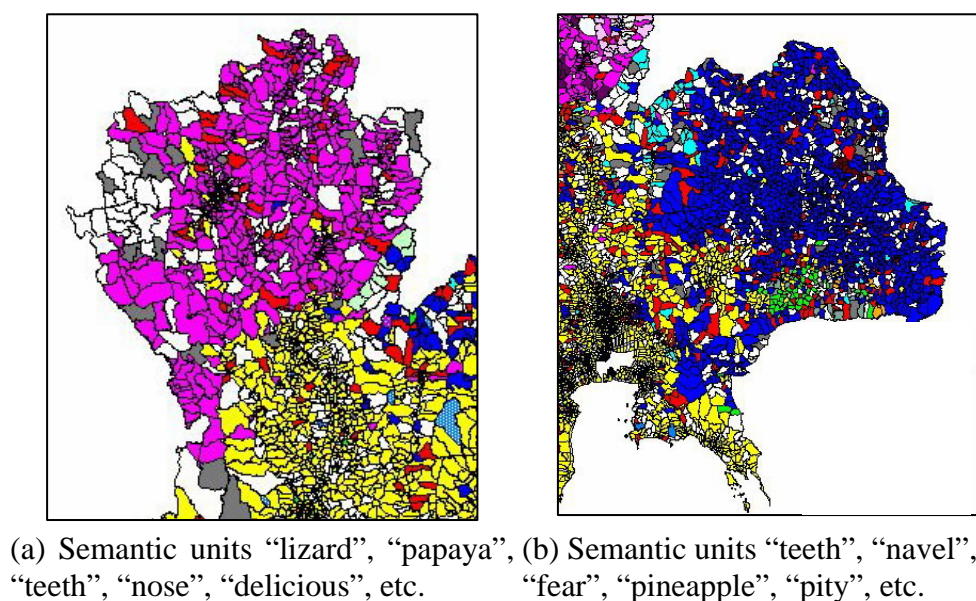
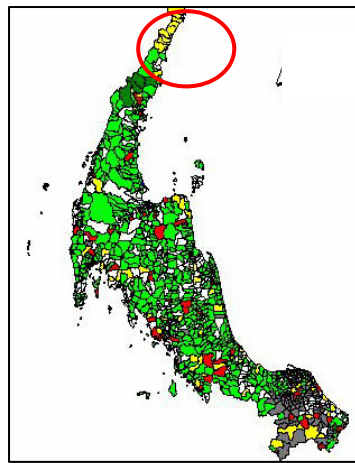


Figure 4: Map showing clear and stable boundary (a), and (b)

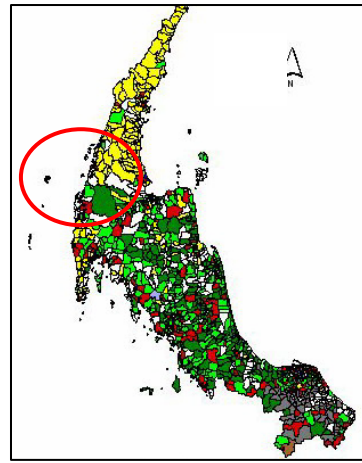
## Conclusion and further work

Results regarding the construction of lexical GIS database and the 170 lexical variation maps have shown the potential of GIS being applied for this project, especially in terms of the ease of map display, error checking, and lexical data retrieval. Further analysis of the lexical data is required especially the regrouping of the multi-item locations which appear mostly in the Northeastern region inhabited by many ethnic minority groups. Reanalysis of the items that vary phonologically but are not dialect markers and also cross-analysis among the semantic units are on the agenda.



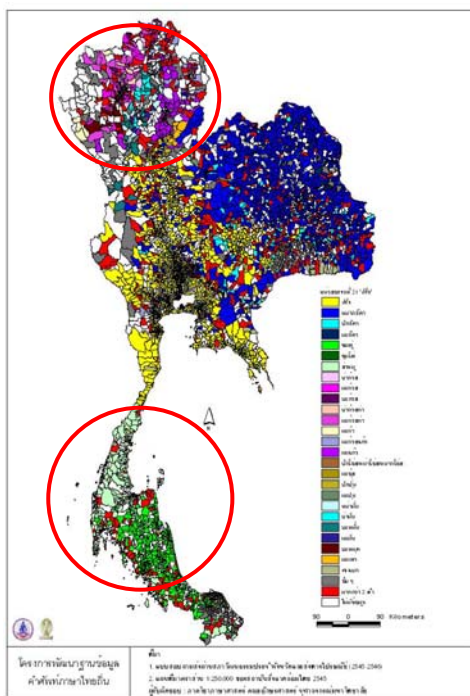


(a) Semantic unit “fun”

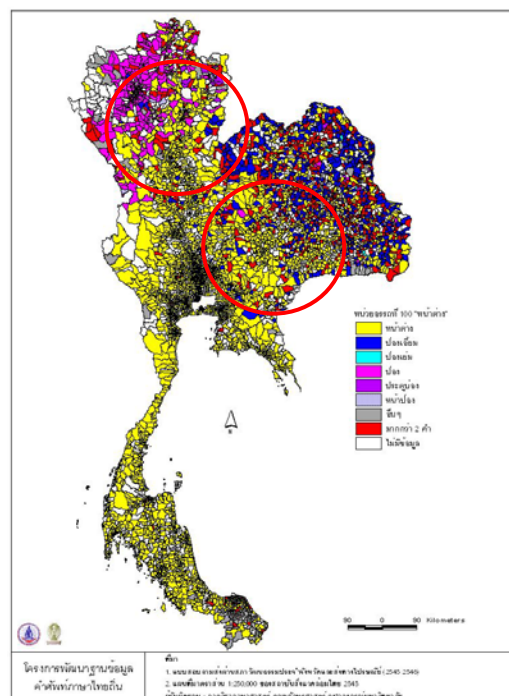


(b) Semantic unit “to speak”

Figure 5: A lexical variation map of the semantic units (a) “fun” and (b) “to speak”.



(a) Semantic unit “guava”



(b) Semantic unit “window”

Figure 6: (a) map showing subdialect areas, and (b) map of “window” showing the spread of the word /naa<sup>T</sup> taaN<sup>T</sup>/ which refers to the modern styled window at the expense of /pON<sup>T</sup>/ and /pON<sup>T</sup> ?iam<sup>T</sup>/ which refer to the traditional styled window.

## **Reference:**

- Chalida Rojanavathanavuthi (1990). ภูมิศาสตร์คำศัพท์ภาษาไทย 4 ถิ่น : การศึกษาเชิงจำกัด (Word Geography of the Four Thai Dialects: a Selected Study). Research report submitted to Chulalongkorn University.
- Duangjai Ache (1985). ภูมิศาสตร์คำศัพท์ภาษาไทยถิ่นใต้จังหวัดสุราษฎร์ธานีและนครศรีธรรมราช (Word Geography of Southern Thai Spoken in Surat Thani and Nakhon Si Thammarat). M.A. thesis, Department of Linguistics, Graduate School, Chulalongkorn University.
- Heywood, I., S. Cornelieus and S. Carver (2002). An Introduction to Geographical Information Systems. New York, Prentice Hall.
- Jaroon Bunpan and Wantanee Pantachat (1983) “การศึกษาการกระจายศัพท์ในประเทศไทย” (A Study of Lexical Distribution in Thailand). Science of Language Papers 3, pp. 27-62. Faculty of Arts, Chulalongkorn University.
- Kalaya Tingsabadh (2001) “Thai Tone Geography”. In Essays in Tai Linguistics, pp 205-228, edited by Kalaya Tingsabadh and Arthur S. Abramson. Bangkok: Chulalongkorn University Press.
- Longley, P. A., M. F. Goodchild, D. J. Maguire and D. W. Rhind (2005). Geographic Information Systems and Science. West Sussex, John Wiley & Sons Ltd.
- Vichintana Panupong (1986). ภูมิศาสตร์คำศัพท์ในจังหวัดนครราชสีมา: โครงการนำร่อง (Word Geography of Changwat Nakhon Ratchasima: a Pilot Study) Research report. Bangkok: Chulalongkorn University Printing House.
- Wantanee Pantachat (1983). ภาษาถิ่นย่อยของคำเมือง: การศึกษาคำศัพท์ (Regional Varieties of Khammuang: a Lexical Study), M.A. thesis, Department of Linguistics, Graduate School, Chulalongkorn University.

---

## **Acknowledgement**

The authors wish to thank the following agencies and individuals for their support of this project. The Office of the National Culture Commission, Ministry of Culture, gave us financial support as well as allowed us to use their network for data collection. The officials of the ministry in every part of the country helped distribute and collect the questionnaire. Faculty of Arts, Chulalongkorn University, allowed us to use the computers in the library computer laboratory for data input free of charge for over a year. Numerous assistants spent long hours with data input. Finally, the Thailand Environment Institute allowed us to use their Thailand map. Without such support this project would not have been completed.

## **ABSTRACT**

### **Word Geography of Thailand**

M.R. Kalaya Tingsabadh\*, Phinnarat Akharawatthanakun, Uraiporn Tantinimitkul, Sirivilai Teerarojanarat, Ornusa Chinkrachangkit

Faculty of Arts, Chulalongkorn University,  
Phyathai Road, Bangkok, Thailand  
kalaya.t@chula.ac.th

Most of the previous word geography studies of Thai dialects cover either a region or a few provinces. In 1996 the word geography of Thailand project was launched under the sponsorship of the Ministry of Culture. A questionnaire of 170 semantic units was constructed. These units all show interesting geographical distribution of lexical items in previous studies i.e. the boundaries between the four main dialects and the existence of subdialect and local dialect areas. The purpose of this study is to see the extent of variation in these semantic units in the context of the whole country.

The questionnaire was sent initially via the Ministry of Culture network and then by post to all of the tambon, the smallest administrative unit, in Thailand. The capital city, Bangkok, was excluded due to its cosmopolitan environment. Out of the 7,255 questionnaires sent, 6,292 or 86.7% came back and were used in this project.

The data were compiled using Microsoft Access and Excel. The items that differ phonologically were analyzed as one lexical item except in the cases where there might be dialect or subdialect differentiation. ArcView, a GIS software, was used to produce the 170 maps - one map per semantic unit.

Results show that the Northern-Central Thai and the Northeastern-Central Thai boundaries are quite clear and stable, whereas the Central-Southern Thai one is not. Some of the maps show the areas of subdialects and local dialects. At least one map shows how lexical change comes with new lifestyle.

Further analysis of the data is required especially the regrouping of the multi-item locations which appear mostly in the Northeastern region inhabited by many ethnic minority groups. Reanalysis of the items that vary phonologically but are not dialect markers and also cross analysis among the semantic units will be carried out.