

International Conference: Spatial Thinking and Geographic Information Sciences 2011

Using GIS for Linguistic Study: a Case of Dialect Change in the Northeastern Region of Thailand

Sirivilai Teerarojanarat^{a*}, Kalaya Tingsabadh^b,

^a*Department of Geography, Faculty of Arts, Chulalongkorn University, 10330, Thailand*

^b*Department of Linguistics, Faculty of Arts, Chulalongkorn University, 10330, Thailand*

Abstract

While spatial thinking is the main concern for all disciplines involving the study of distributed phenomenon in space and over time, lack of spatial abilities and knowledge in handling spatial data appears to be one of the weakest points of researchers working in the area of social studies and humanities. This paper aims to present a way by which linguistic approaches and spatial-based techniques using a GIS tool are integrated to help spatially investigate and quantify the dialect change at two different points of time. Applied methodology with the exemplified application to the whole northeastern region of Thailand is illustrated and discussed. Results are promising and make a genuine contribution to the field of linguistics. The role of GIS is obvious, showing its great potential in advancing our fundamental understanding of the spatial pattern of dialect change as well as its relation to other related factors such as social and cultural changes.

© 2011 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).

Selection and/or peer-review under responsibility of Yasushi Asami

Keywords: Dialect change; GIS; Thailand;

1. Introduction

Without doubt, almost all aspects of science and disciplines nowadays are increasing in demand for high-quality information so as to realistically model the real world and make decisions from reliable information. Most information, to some degree, unavoidably involves a spatial component and requires ways of handling it spatially - measuring, mapping, analyzing or displaying it with respect to location on the earth's surface.

Linguistics is no exception. Its relevant sub-branch is dialectology - the study of regional variation. Subject matters to be investigated can be syntax, vocabulary, consonants, vowels, or tones. The methodology used in this discipline involves field survey, data collection and recording, and map making. Dialectologists are interested in locating where dialects are spoken involving specifically identification of

* Corresponding author. Tel.: +662-2184821; fax: +662-2184822.

E-mail address: sirivilai.t@chula.ac.th.

boundary positions. Research on dialect boundaries has shown that features of one dialect may spread into the area of a neighboring dialect as a result of external factors such as social and cultural influence.

Change is also of much interest. According to Langacker [1], “Every language is the product of change and continues to change as long as it is spoken”. This is obviously true as we can simply notice that new words are regularly added due to the innovation of new concepts, materials, or technologies, etc. For example, words like “web page” or “internet” are recently invented and used worldwide nowadays. On the other hand, some words have become less frequently used, or obsolete, or replaced by other words in more prestigious languages. Studying dialect change can, thus, provide us with an opportunity to explore and examine the pattern of change – its direction and, possibly, its rate of change. Historical study of locality can also gain benefit by analysing the traces of dialect change and language contact. As dialects are a significant part of our culture, the study of dialect change is necessity, not only for the interest of linguists but as an important part of the study of human cultural heritage and the historical evidence of mankind.

Up to now progress in dialectology has been much hindered by the limitation of spatial abilities and background knowledge in handling spatial information. Points of data collection have always been roughly marked. Map making and cartographic display have usually been created manually on paper maps. Usually, overlay technique i.e. for the comparison of dialect change has been done manually by superimposing a bundle of paper maps using different scales of map sources and so on. These limitations, to some degree, impact the accuracy and quality of data recorded and analysis which will probably result in the creation of unreliable map results.

The emergence of the Geographic Information System (GIS) in the 1960s [2] has been proven a very powerful tool to handle spatial data, mainly for data storage, database management, spatial analysis, and cartographic display. Details of its capabilities are given extensively elsewhere, e.g., the textbooks of Tomlin [3], Maguire et al. [4], Longley et al. [5]. However, the applications of GIS to linguistics have been done just over the last few decades, e.g., the work of Lee and Kretschmar [6], Luo et al. [7], Wang et al. [8] and Ayad and Luthin [9], etc.

In 2009, the GIS-based Linguistic Geography of Thailand Project was initiated under the sponsorship of Chulalongkorn University with its key aim to promote the use of Geographic Information System (GIS) in linguistics. The project enables scholars from different fields of knowledge to work together, in this case, geographer and linguist. A series of research work has been conducted since then. The first series was the research work of the Word Geography Maps of Thailand project, producing a geographic database of 170 Thai dialect vocabularies based on the data collection in 2002-2003 [10]. The second series was the extension work of the first project - the creation of the boundary map of Central and Non-Central Thai Dialects by overlaying 170 map layers [11]. A recent work was the Word Geography Maps of the Northeastern Thai Dialect. In this recent work, the paper maps of Thai dialect vocabularies in the northeastern region of Thailand based on data collection in 1979 was converted and transformed under a GIS environment to be available in a digital map. Visit these projects via the website the GIS-based Linguistic Geography of Thailand (<http://ling.arts.chula.ac.th/geoling/>).

The overlap study area but with different dates in the 1979 and 2002 studies, which is approximately 20 years apart, allows the dimension of time to be investigated. This paper thus serves as an example for demonstration of an applied methodology by which the traditional linguistic approach is integrated with GIS to enhance the way to spatially compare and investigate the change of dialect vocabularies at two different points of time. The applied methodology, results and discussion as well as the benefit of GIS for the study are also given.

2. Study area and scope of study

Location of the study area is the whole northeastern region of Thailand, covering an area of about 170,000 sq. km. This area is namely the Northeast or Isan as Thais call it. Based on political

administrative boundaries, the Northeast nowadays consists of 17 provinces. It borders Laos to the north and the east, Cambodia to the south, and the central region of the country to the west.

Geographically this part of the country, lying in the heart of the Korat Plateau, is a region dominated by a rolling surface and undulating hills and irrigated by the Shee and Mun River. Because of the topographic difference, the Korat Plateau of the Northeast can be visibly separated from the flat area of the Central Plain and the mountain ranges of the North. In terms of language, Thai is the official national language spoken in every part of the country. Four main Thai dialects are Northern Thai, Northeastern Thai, Central Thai, and Southern Thai. Their names are differentiated by the parts of the country where these dialects are spoken. Northeastern Thai is the major dialect in the Northeast. In this study where dialect change will be investigated, spatial variation of Central Thai and Northeastern Thai in the area of the study is the main focus.

3. Methodology and analysis

According to the linguistic approach, studying the language change can be conducted by two techniques: real time and apparent time. Each technique has a variety of pros and cons which has been detailed and discussed elsewhere [1, 12]. According to the first technique, language data collection and observation need to be made at the same community at a series of times e.g. at two different points in time. With this technique, it is very inconvenient as we have to wait, probably twenty years, to see the change. A study based on this technique is up-to-now quite rare but it is realistic as it is performed at the actual time of study. Another technique of study, the apparent time, is usually used as an alternative because it is more flexible to perform. Based on this technique, data collection from at least two age groups e.g. the elders and the younger generation at a particular community is compared to detect changes. Most studies have been carried out using this technique.

Over the past decades the apparent time technique was mostly employed in the study of dialect change in Thailand. This study which aims at detecting dialect change is different from the previous studies in two aspects. Firstly, as it is fortunate to have two studies carried out in the same area approximately 20 years apart, the technique of real time investigation can be applied. Secondly, unlike those studies, the spatial-based technique using a GIS tool was integrated into the conventional linguistic approach in this study. To this end, GIS will play a key role in helping to explore spatially the pattern of dialect change - its direction and its rate of change.

Based on the two main sources of language data as briefly described in the introduction, analysis of the study was divided into four main steps. The first step involves analysis and classification of the lexical items or words of each semantic unit into groups based on linguistic similarity. In this study, dialect data was classified into four classes. These classes were used commonly for both datasets. They are (1) Central Thai, (2) Northeastern Thai, (3) a mixture of more than one dialect, and (4) other dialects. The second step is to produce isogloss maps. An isogloss is a line drawn on a map dividing two regions which differ with respect to, in this case, a lexical item. That is, each semantic unit, classified as one of the four classes from the previous step, was coded and shaded as an isogloss map in GIS. The third step was to investigate the datasets of two different dates to be spatially comparable. Fig 1. shows the data collection units of the two datasets in comparison. Fig 1(a) was the first dataset of the study - the Word Geography Maps of the Northeastern Thai Dialect project. It used the grid cell (of about 25 km x 25 km) whose center was used as collection unit. The second dataset of the study (see Fig 1(b)), from the Word Geography Maps of Thailand project, used the whole area of sub-district, or “*tambon*” in Thai, as collection unit. In summary, due to the different techniques of data collection, the first dataset was stored in a GIS database as point and polygon features while the second dataset was stored as polygon features. As shown in Fig 1(c), once superimposed, it is obvious that although both datasets cover the same geographically-referenced area and can be overlaid concisely, their scale and spatial resolution of data collection units were totally different.

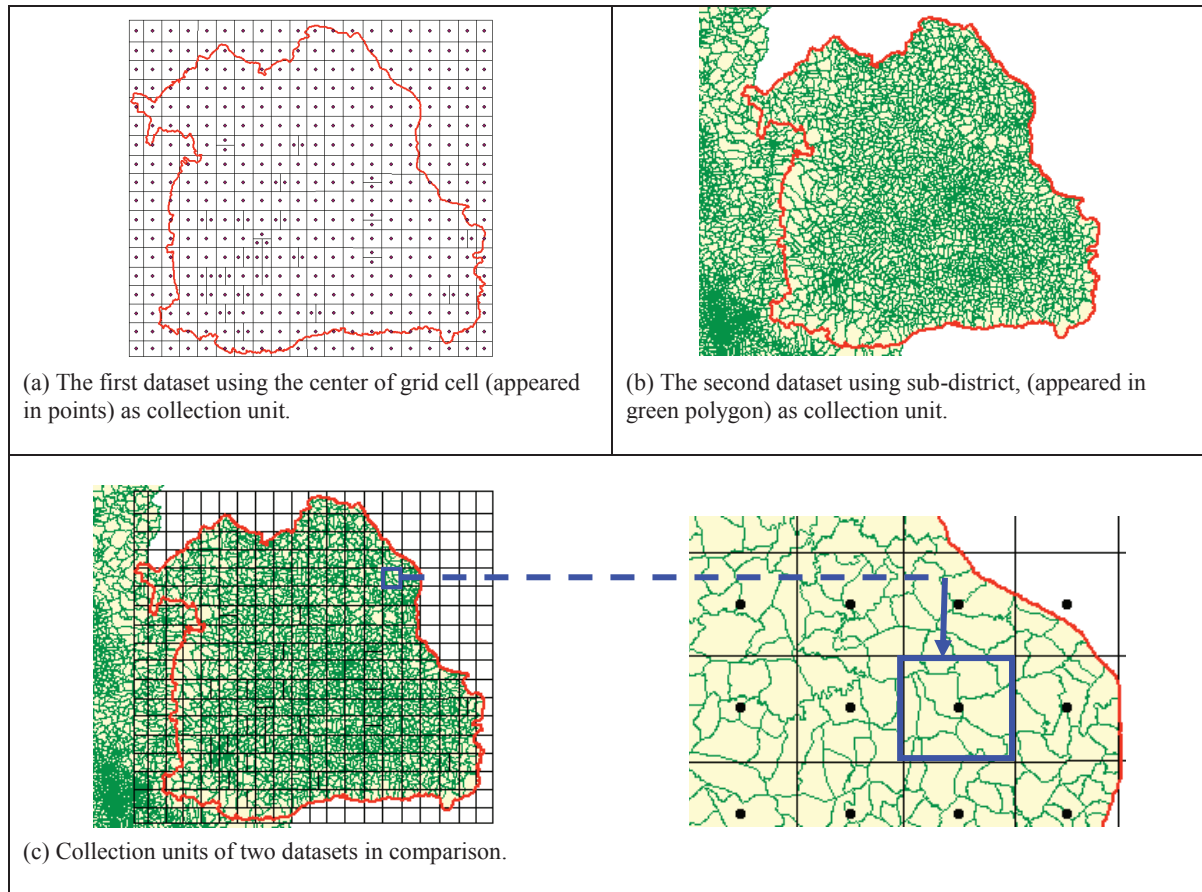


Fig. 1. Data collection units of the two datasets used in the study area in comparison.

According to the investigation, the initial agreement was made beforehand that the spatial comparison between the two datasets will be performed at some parts of tambons (as polygons) where each center of the grid cell (or centroid, as called in its technical term) within a radius of 5 Kilometers falls only as shown in Fig 2. Once the agreement was made, the last step was to perform the spatial comparison. In this study, the spatial comparison was made in two ways: visual observation and quantification of the dialect change. By means of visual observation, a pair of the same semantic unit of different dates was simply superimposed as shown in Fig 3(a). From the Figure, point features represent dialect in 1979 and shaded polygon features represent dialect in 2002. This way the overall picture of dialect change can be observed. Another way was to quantify the change. Based on the previous agreement, the spatial overlay analysis on the basis of ‘intersect’ operation was performed for each pair of the same semantic unit of different dates in order to create a composite map of dialect change patterns. The new composite map produced contains both sets of dialect attributes of the two different dates. Quantification of change, thus, can be extracted from these attributes and displayed as the area of change as shown in Fig 3(b). From the Figure, circles with different colors refer to the intersected area of the two different dates where no change or changes were found. In the next section, results based on the applied methodology are reported.

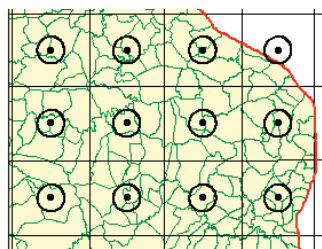


Fig. 2. The initial agreement for spatial comparison in the study.

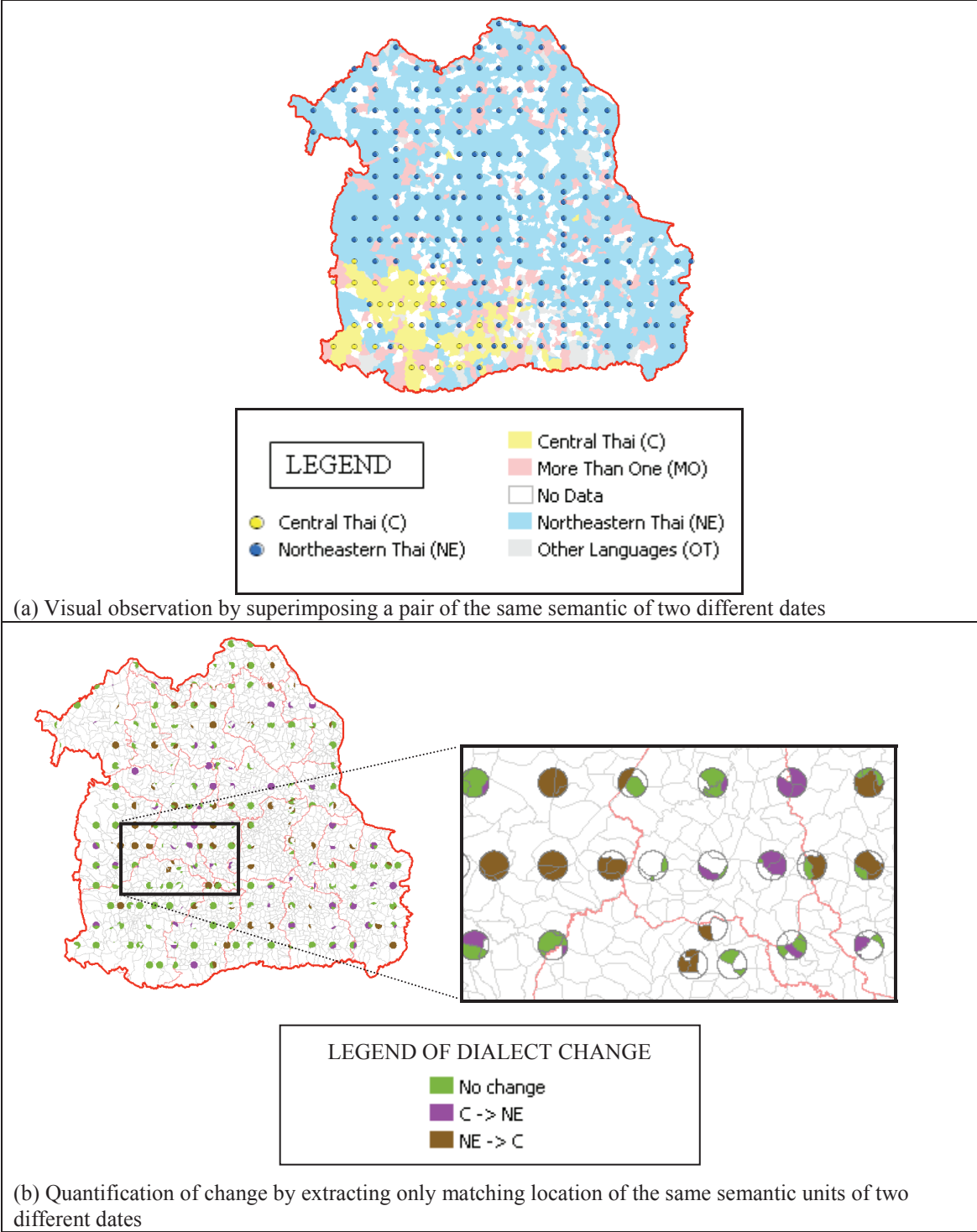
4. Results and discussion

Out of the 298 lexical variation maps in the first data source and the 170 lexical variation maps in the overlap area in the second data source, only 111 semantic units are shared and can be used for comparative analysis. However, in this paper three semantic units are reported. These semantic units are ‘barn’, ‘fan’, and ‘window’. Their selection was based on the fact that these semantic units have been observed to have different degrees of change from Northeastern Thai to Central Thai lexical items [13, 14]. For the semantic unit ‘barn’, people in the Northeast tend to use the Northeastern Thai lexical item most frequently. For the semantic unit ‘window’, people in the Northeast tend to use the Northeastern Thai lexical item least and prefer to use the equivalent Central Thai lexical item. In other words, the Northeastern Thai lexical item of the semantic unit ‘window’ is likely to disappear the fastest. This leaves the semantic unit ‘fan’ in the middle degree of loss or disappearance.

On the basis of visual observation, Fig 4 illustrates the superimposing of the two different dates for semantic unit ‘barn’, ‘fan’, and ‘window’, respectively. When observing these three semantic units independently, the following can be concluded. In the semantic unit “barn” the Northeastern Thai lexical item was used in most areas and the Central Thai lexical item was used in Nakhon Ratchasima province. A similar pattern was found in the semantic unit “barn” but the Central Thai lexical item was found in a larger area where Chaiyapum, Nakhon Ratchasima, and Buriram province locate. In the semantic unit “window” however, the Central Thai lexical item was used in most areas.

Apart from visual observation, quantifying dialect change was performed in two different ways; map of dialect change (see Fig 5) and table of dialect change (see Table 1). It should be noted that only the change between Central Thai and Northeastern Thai was focused on in this paper. According to Fig 5 and Table 1, the abbreviation of C->NE stands for the change from Central Thai to Northeastern Thai, NE->C stands for the change from Northeastern Thai to Central Thai. The symbolic C->C stands for no change of Central Thai between 1979 and 2002, and NE->NE is for the no change of Northeastern Thai between 1979 to Northeastern Thai in 2002. That is, in both cases of C->C and NE->NE, no change was found between the two different dates.

Results obtained from the quantifying dialect change method in the three semantic units are as follows. In the semantic unit “barn” the change between Central Thai and Northeastern Thai was minimal and equal. In other words, no change was dominant, abbreviated as C->C or NE->NE. All changes were found in the southwest of the study area as shown in Fig 5(a). In the semantic unit “fan” the dialect change was more than that of the semantic unit ‘barn’ (Fig 5(b)). In the semantic unit “window” the dialect change was found most with dispersion pattern (see Fig 5(c)). Moreover, the change from Central Thai to Northeastern Thai in this semantic unit was higher than the other two cases. Based on the quantification used, the finding is that the rate or degree of dialect change differs from one semantic unit to another. While the change was hardly found in the semantic unit ‘barn’, it clearly occurred in the semantic unit ‘window’.



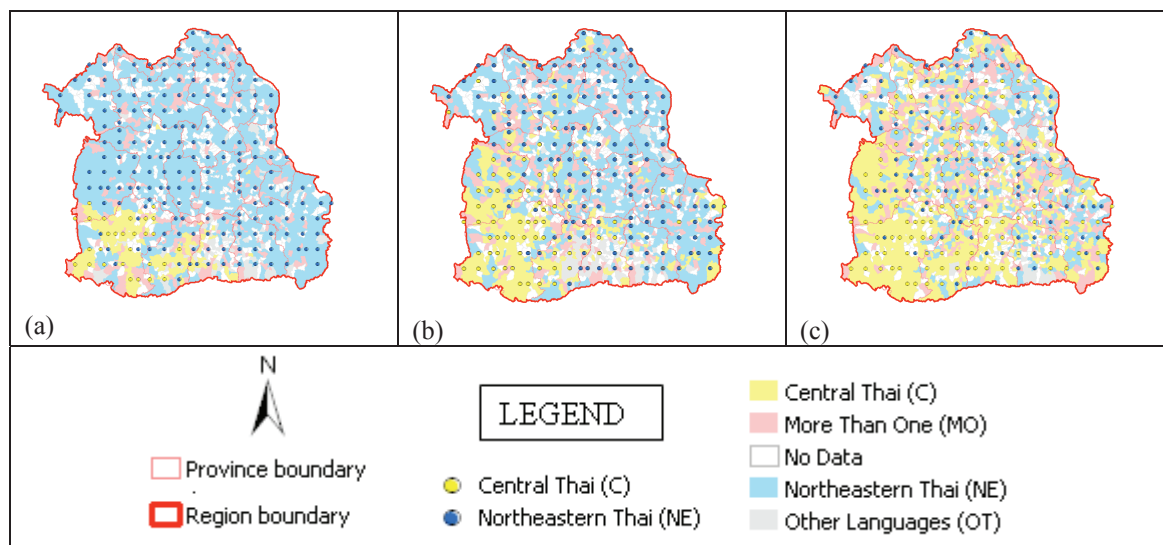


Fig. 4. Superimposing of two different dates of semantic unit (a) ‘barn’, (b) ‘fan’, and (c) ‘window’ in comparison.

Table 1. Comparison of dialect change amongst the 3 semantic units.

| Status of Change | semantic unit ‘barn’ Percent of change | semantic unit ‘fan’ Percent of change | semantic unit ‘window’ Percent of change |
|------------------|---|--|---|
| C->C | 8.46 | 17.31 | 34.54 |
| NE->NE | 84.29 | 62.94 | 24.67 |
| C->NE | 2.27 | 9.22 | 17.60 |
| NE->C | 4.98 | 10.52 | 23.19 |
| Total | 100.00 | 100.00 | 100.00 |

Remark: “C” stands for Central Thai and “NE” for Northeastern Thai.

All in all, we can conclude that the pattern of dialect change of the semantic units investigated in this paper, both in terms of their direction and their rate of change, is different from one to another. The dissimilarities of these change patterns can be explained as the unique usage of each single word. This finding is in line with the saying that “every word has its own history”.

On the basis of the findings, one could also predict the future of each Northeastern Thai lexical item investigated. While the Northeastern Thai lexical item for “window” may soon disappear, that for “barn” will be retained long into the future. Comparing the retention rate of the Northeastern lexical items in the 111 semantic units will yield interesting results. It is also expected that different sets of semantic units, e.g., utensils, animals, plants, and kinship terms, would not have the same retention rate. Classifying the semantic units into such groups and comparing the retention rate across groups would likely provide insight into some factors influencing dialect change. Furthermore, calculating the total retention rate at each location and comparing the rates across locations could provide background information for further studies - social, cultural and historical. These analyses will be carried out and reported in the future.

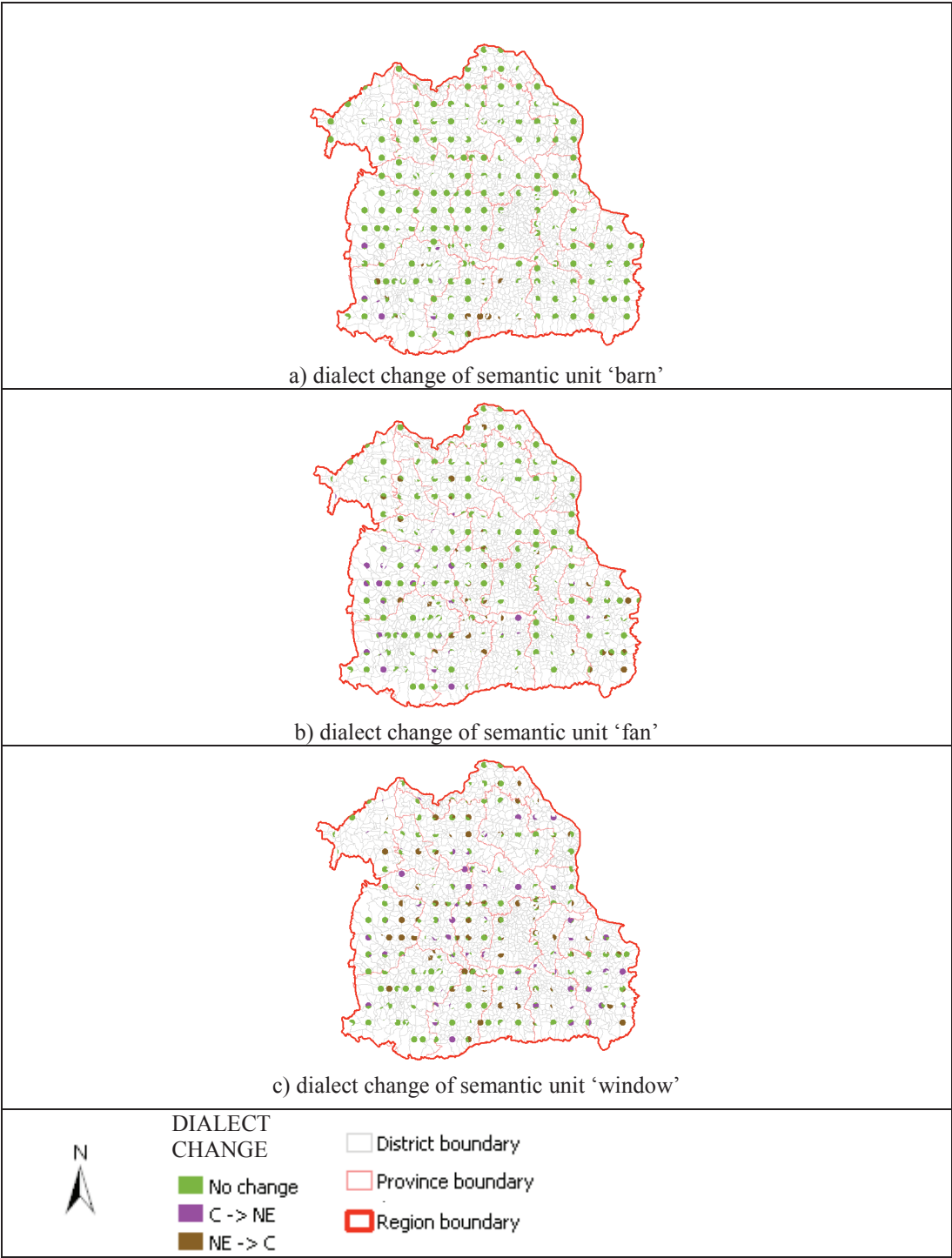


Fig. 5. Spatial pattern of dialect change of three semantic units.

5. Conclusion and recommendation

Owing to the fact that all dialects gradually change over time, this paper presents an alternative by which the spatial-based technique under a GIS environment was applied to the linguistic study of dialect change. Based on the availability of two different dates of previous works, the technique of real time was exploited. Some specific semantic units were investigated to plot a spatial pattern of dialect and its change direction and rate. In this study GIS facilitates the spatial tools and functions, mainly for cartographic display and quantification of dialect change. Although simple overlay and query function were applied, the contribution to linguistics is significant. The applied methodology benefits linguists and researchers in many ways. Firstly, GIS can produce different views of maps. Researchers can then learn and get better knowledge on seeing the spatial pattern of lexical variation/classification and the spatial shift of dialects. Secondly, GIS can produce a more accurate overlay map. Overlaying a series of maps with different data sources, scales, or spatial resolutions can be handled efficiently. Thirdly, GIS can produce a quantified overlay map and tabular report as shown in the result section. Fourthly, the repetitive task such as performing a spatial overlay for each pair of 111 semantic units can be conducted conveniently by writing programming scripts or using customized tools such as model builder of ARCGIS software. In addition, the resultant maps can ease the superimposing of other types of available data such as population data and settlement.

To conclude, GIS has shown its great potential in advancing our fundamental understanding of the spatial pattern of dialect and its change. In our case, further analysis and questions were left to linguists to find the reasons behind these diversified changes. That is, further study of the related interplay between social and cultural changes, history, migration, physical environment, innovation, as well as advancement in communication and transportation etc. that lies behind the spatial pattern of dialect and its change should be investigated.

Acknowledgements

The authors wish to thank Chulalongkorn University Centenary Academic Development Project 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

References

- [1] Langacker RW. *Language and its structure: some fundamental linguistic concepts*. Harcourt, Brace & World: New York; 1968, p.175
- [2] Goodchild M F. Geographical information science. *International Journal of Geographical Information Systems* 1992;6(1)31-45.
- [3] Tomlin C D. *Geographic Information Systems and Cartographic Modelling*. Prentice Hall: Englewood Cliffs; 1990.
- [4] Maguire D J, Goodchild M F, Rhind D W. *Geographical Information Systems: Principles and Applications*. Longman Scientific and Technical: Harlow; 1991.
- [5] Longley P A, Goodchild M F, Maguire D J, Rhind D W. *Geographic Information Systems and Science*. John Wiley & Sons: West Sussex; 2005.
- [6] Lee J and Kretzschmar W A. Spatial analysis of linguistic data with GIS functions. *International Journal of Geographic Information Systems* 1993;7(6):541-560.
- [7] Luo W, Hartmann J F, Li J, Sysamouth V. GIS Mapping and Analysis of Tai Linguistic and Settlement Patterns in South China. *Geographic Information Sciences* 2000;6(2):129–136.
- [8] Wang F, Hartmann J, Luo W, Huang P. GIS-Based Spatial Analysis of Tai Place Names in Southern China: An Exploratory Study of Methodology. *Geographic Information Sciences* 2006;12(1):1-9.
- [9] Ayad Y, Luthin H. Mapping Dialect: GIS Applications in Studying Dialect of Western Pennsylvania, In *The 2009 ESRI International User Conference Proceedings*, 13 - 17 July, San Diego, California, USA; 2009, p1-26, Access: http://proceedings.esri.com/library/userconf/proc09/uc/papers/pap_1686.pdf.

- [10] Teerarojanarat S, Tingsabadh K. Development of Geographical Database of Lexical items in Thai Dialects. *Journal of Language and Linguistics*. Department of Linguistics, Faculty of Liberal Arts, Thammasat University 2008; **26**(2):1–20. (in Thai).
- [11] Teerarojanarat S and Tingsabadh K. A GIS-based Approach for Dialect Boundary Studies. *Dialectologia*. 2011; **6**:53-73. <http://www.publicacions.ub.edu/revistes/dialectologia6/>.
- [12] Chambers J K and Trudgill P. *Dialectology*. Cambridge University Press: Cambridge; 1980.
- [13] Panupong V. *Word Geography of Nakhon Ratchasima : A Pilot Project*. Chulalongkorn University Printing House: Bangkok; 1986. (in Thai).
- [14] Rinprom C. *Studies in Thai Dialects*. PhD dissertation, University of Copenhagen, Copenhagen, Denmark; 1987.