Potential Applications of GIS for Linguistic Data

メタデータ	言語: eng 出版者: 公開日: 2022-04-06 キーワード (Ja): キーワード (En): 作成者: H. Lowry, John メールアドレス: 所属:
URL	https://doi.org/10.15021/00009901

Senri Ethnological Studies 108: 59–74 ©2022
Fijian Languages, Cultures, and Their Representation
Edited by Kikusawa Ritsuko and Sano Fumiya

5. Potential Applications of GIS for Linguistic Data

John H. Lowry
Massey University, New Zealand

Abstract

The aim of this chapter is to briefly outline potential applications of GIS for linguistic data. As a computer database, GIS not only offers an efficient means for managing large amounts of complex spatial data but also provides a variety of tools for analysing and visualizing spatial patterns and relationships. Linguistic geography, or geolinguistics, is an interdisciplinary field that recognizes the importance of spatial patterns and geographic relationships as drivers of change and the diversity of language. The chapter provides an overview of case studies where linguistic data were used in a geospatial context. Two case studies are characterised as dialectometric studies with the aim to measure variation in dialects in relation to geography and geographic distance. The third case study examines the role of physical environmental features on linguistic variation of toponyms. Following this overview of case studies, the chapter introduces the Fijian Language GIS Project, which is an interdisciplinary effort with goals to: i) develop a GIS database of Fijian communalects, ii) use the database to conduct scholarly research on linguistic variations in Fiji from a spatial perspective, and iii) produced information suitable for dissemination to the public through museums and other venues. The chapter provides a brief description of the GIS database of Fijian communalects and how it was created followed by an overview of a dialectometric analysis of the data for a pilot study area on the island of Kadavu. The chapter concludes by discussing some of the opportunities and challenges the project faces. These include the production of a rich database of Fijian communalects that will be of interest to scholars and lay persons alike, but with challenges associated with efficiently disseminating and meaningfully analysing a rather large and complex database.

5.1. Background

Both cultural geographers and linguists recognize the value of mapping linguistic information. While cultural geographers tend to focus on the spatial differentiation of language as an artefact of the cultural landscape (Jordan 2014), linguists tend to be more interested in the evolution of language and the complexities of language variation (Haynie 2014). Linguistic geography, or geolinguistics, is an interdisciplinary field that recognizes the importance of spatial patterns and geographic relationships as drivers of change leading to the structural diversity of language (Haynie 2014). With roots in traditional core ideas of cultural geography, such as diffusion theory (e.g. wave model (Hägerstrand 1966) and gravity model (Olsson 1965)), the field of geolinguistics has benefited in recent decades from advances in computational analysis (e.g. data mining) and Geographic Information Systems (GIS). As a computer database, GIS not only offers an efficient means for managing large amounts of complex spatial data but also provides a variety of spatial and statistical tools for analysing and visualizing spatial patterns and relationships. Several authors have noted the potential of GIS to facilitate the analysis of linguistic data based on common geographic concepts such as spatial scale, proximity, spatial autocorrelation, and geographic distance (Haynie 2014; Hoch and Hayes 2010).

An important role of GIS in linguistic geography lies in the exploration of innovative approaches for conveying complex geolinguistic phenomena through cartographic display and visualisation. Challenges associated with mapping linguistic information include geographic scale, placement of boundaries, and size of map units (Luebbering 2013). The spatial scale of language patterns correlates with the amount of time languages have had to evolve. Patterns of micro-variation in dialects occur at a regional scale reflecting a relatively recent stratum of history, whereas global patterns of divergent languages represent deeper histories (Haynie 2014). With regards to the placement of boundaries, linguists and cultural geographers are well aware of the fact that an isogloss represents an imaginary division between dialects or languages that is neither precise, nor oftentimes, objective (Haynie 2014). However it must be remembered that maps are powerful tools capable of intentionally or unintentionally misleading the map reader (Luebbering 2013; Monmonier 2005). The problem of map units is often not one of choice, but of data availability as linguistic data are often collected for political or administrative units that the data user has little control over. Unless mapping occurs at the level of the individual (which is impractical) any geographic unit used for mapping is an aggregation, with associated issues of presumed homogeneity and ecological fallacy. One of the forthcoming challenges in linguistics geography therefore lies in developing innovative and effective means of visualizing geolinguistic data through cartographic display (Luebbering, Kolivras, and Prisley 2013).

Mapping linguistic information through linguistic atlases has been common since at least the middle of the last century (Kurath et al. 1939–1943). With the advent of GIS in the 1980s and the digitization of all types of information, linguistic atlases have become much more accessible. An example of such a digital atlas (in .pdf format) is the *Linguistic Atlas of French Polynesia*, a volume of 2,200 maps documenting the diversity of languages and dialects of French Polynesia (Charpentier et al. 2015) (Figure 5-1). With the emergence of internet-based technologies, maps need not be static or limited to information chosen solely by the map-maker. Internet-based GIS websites provide users with the opportunity to interact with geographically-based linguistic data which greatly increases the amount, and type (e.g. video and audio media) of information accessible to the map reader/user. Examples of internet-based interactive maps include the *First Languages Australia* web map (Baisden et al. n.d.) and the *Algonquian Linguistic Atlas* (Junker 2014–2019).

The aim of this chapter is to briefly outline some applications of GIS for linguistic data.

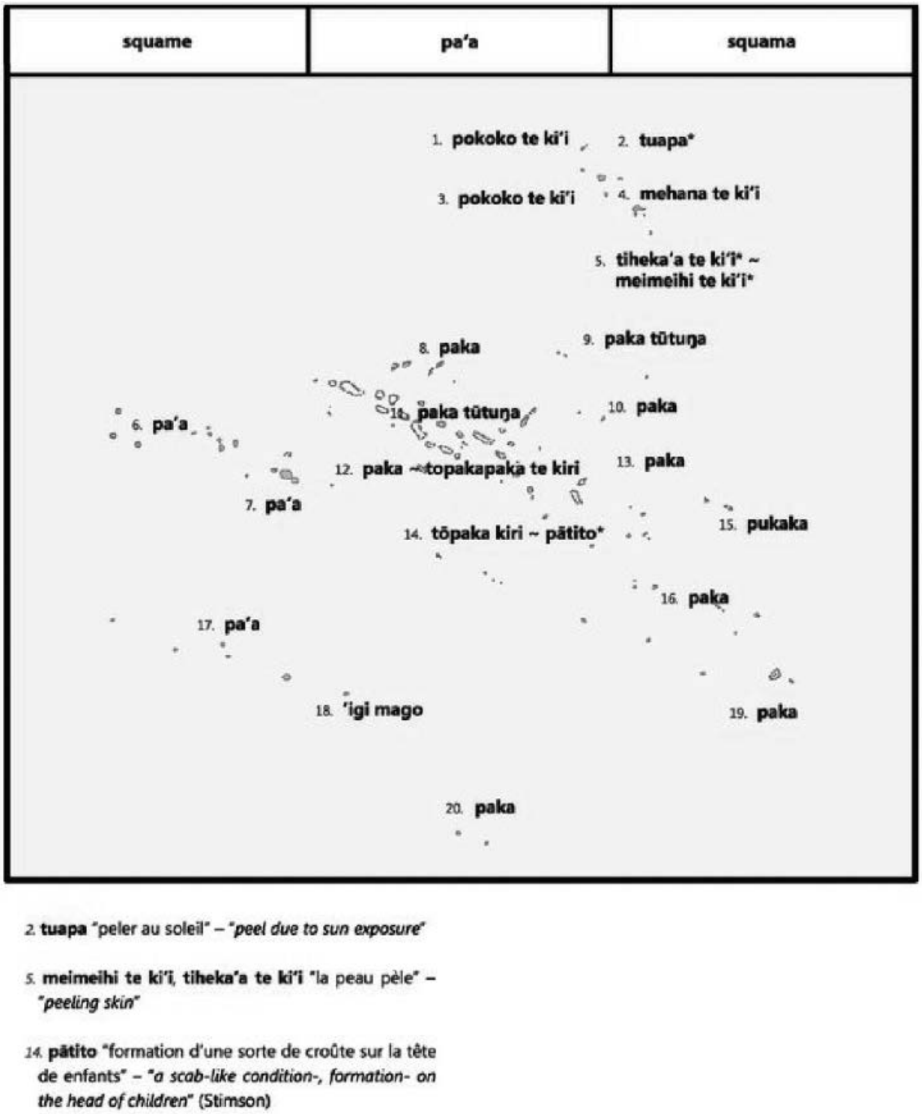


Figure 5-1 Example map from the *Linguistic Atlas of French Polynesia* (Charpentier et al. 2015)

The approach will be to highlight a few case studies where geolinguistic data have been analyzed, particularly in the field of dialectometry. This is followed by an introduction to the Fijian Language GIS Project with an examination of preliminary data analysis and visualizations of a pilot study area. The chapter concludes with a discussion of opportunities and challenges faced by the Fijian Language GIS Project.

5.2. Examples of Spatial Analysis and Cartographic Visualization for Linguistic Data

Dialectometry is a branch of geolinguistics that seeks to measure patterns in large datasets to infer meaning as to the structure of language with an aim to better understand dialect variation and evolution (Haynie 2014; Szmrecsanyi 2011). Given the theoretical relationship between geographic distance and linguistic variation, dialectometry research often seeks to establish correlations between the two (Nerbonne 2013; Nerbonne and Kleiweg 2007; Nerbonne et al. 2008).

Following is a brief overview of two case studies where dialectometric methods with geolinguistic data were used to investigate dialect variations in Switzerland and Great Britain. The final part of this section provides an overview of a case study in southern China that used GIS to analyze linguistic variations in relation to geography, but which would not be considered a dialectometric method.

5.2.1. Dialectometric Methods and GIS

Both case studies examined variations in words and sentence structure (morphosyntax, or grammar) to evaluate differences in regional dialects. Stoeckle (2016) used data from the

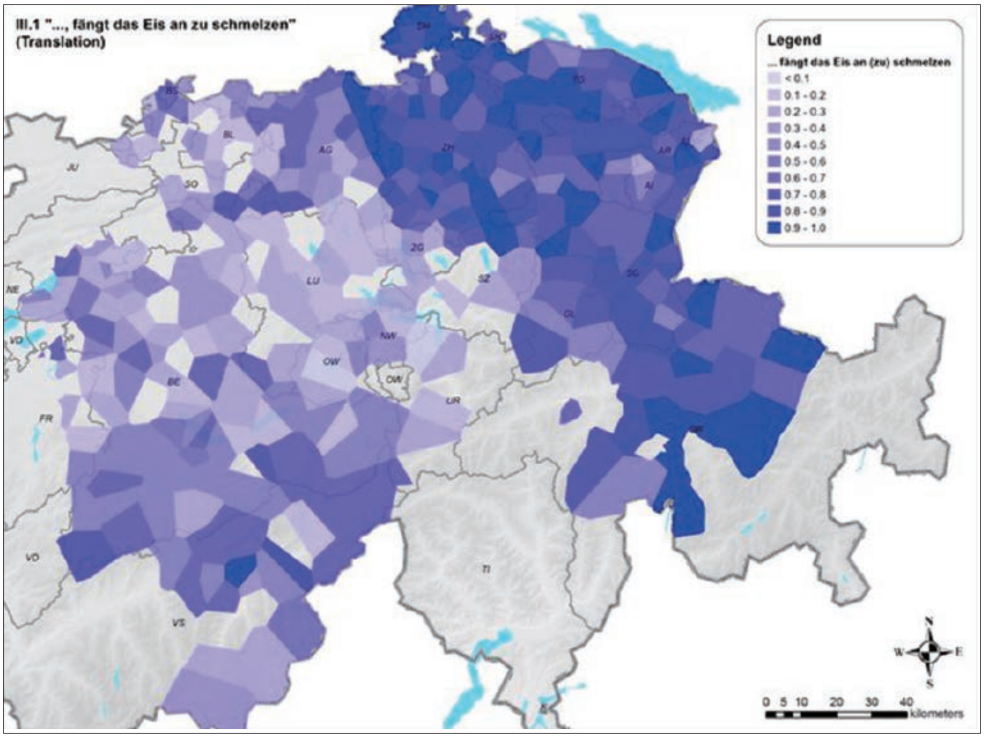


Figure 5-2 An example of the use of GIS to map morphosyntax (1). This map shows the percentage of informants using a particular variant of the sentence *fängt das Eis an (zu) schmelzen*, Point features are extrapolated as area features using Voronoi polygons. (Stoeckle 2016)

Syntactic Atlas of German-speaking Switzerland (SADS) database, which contains the results of a questionnaire survey of 2,770 informants in different locations of German-speaking Switzerland. The aim of the study was to examine regional (i.e. horizontal) variation and interpersonal (i.e. vertical) variation in morphosyntax for a set of 57 sentences, and to visualize the variation cartographically. While the geographic unit of data collection were point features (i.e. villages and towns), Stoeckle used Voronoi polygons to extrapolate the spatial representation to area features improving visualization of dialect variation across space. Figure 5-2 presents a map of the percentage of informants who used a particular variant of the sentence ... *fängt das Eis an (zu) schmelzen* (English: …begins the ice (to) melt). With the SADS database in GIS format, Stoeckle derived several other metrics and indices of morphosyntax to visualize regional variation. These included maps of the geographic distribution of dominant variants (Figure 5-3) and an index of overall variation (Figure 5-4). Stockel's study illustrates the value of geolinguistic data visualization. He initially hypothesized that areas of high vertical variation would be associated with regions of modernity and dynamism. Instead, the results suggested the opposite—areas with higher vertical variation were associated with more traditional, conservative parts of the country.

In the second case study, Szmrecsanyi (2011) used the *Freiburg Corpus of English Dialects*, a major corpus with samples of traditional English dialects for all of Great Britain.

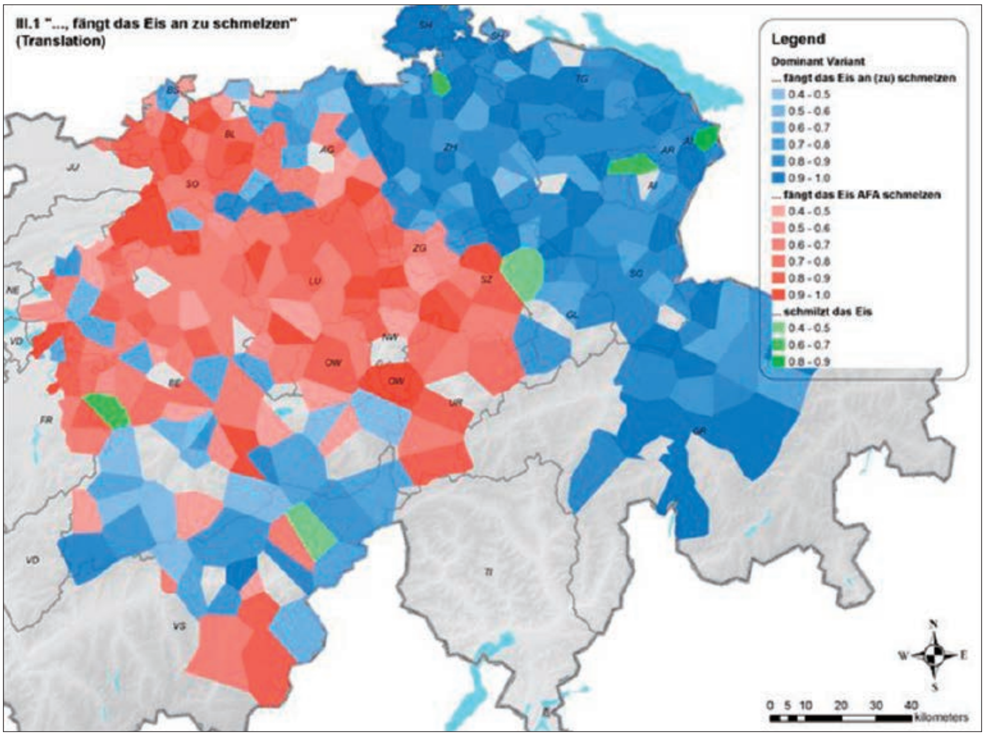


Figure 5-3 An example of the use of GIS to map morphosyntax (2). This is a map of geographic variation of the three most dominant variants of the sentence *fängt das Eis an (zu) schmelzen.* (Stoeckle 2016)

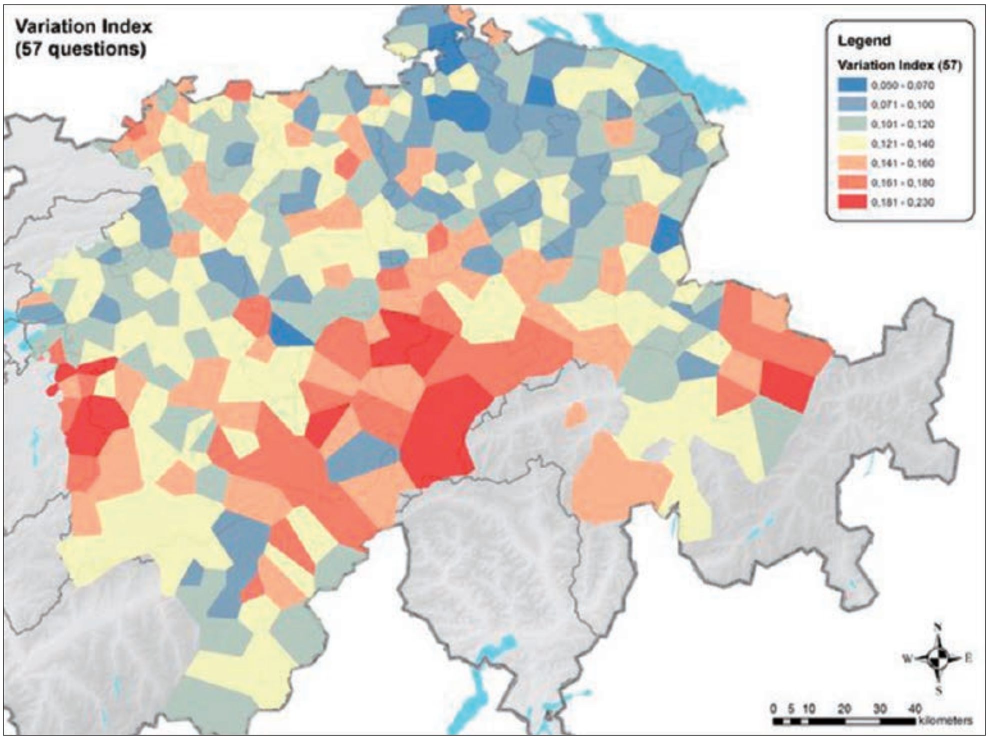


Figure 5-4 An example of the use of GIS to map morphosyntax (3). This map presents an index of overall variation of the 57 variants examined in the study. (Stoeckle 2016)

From the corpus of 368 texts from 431 informants, Szmrecsanyi focused on the frequency of 57 grammatical characteristics (features) found in 34 dialects. To calculate the grammatical distance between the 34 dialects he transformed a 34 x 57 frequency matrix of the features to derive a measure of Euclidean Distance between each dialect. Once the Euclidean Distance matrix was created it was used for visualization and further analysis. Figure 5-5 presents a visualization from the Euclidean Distance matrix showing the location of dialects that are linguistically closer in darker, blueish lines and dialects that are more linguistically distant in proportionally lighter, more yellowish lines. To examine the hypothesis that geographic proximity is associated with linguistic similarity, Szmrecsanyi used a multivariate data reduction technique called Multidimensional Scaling (MDS) to reduce the higher order Euclidean Distance matrix to three arbitrary dimensions. By reducing the distance matrix to three dimensions it was possible to depict linguistic distance along a red-green-blue (RGB) continuum which can be visualized. Figure 5-6 (right) shows a map of linguistic distance for each dialect location using an RGB colour scheme. This can best be interpreted when compared with the left map as a reference point. The left MDS map is based on a geographic (as-the-crow-flies) distance matrix also rendered with an RGB colour scheme. It is clear that dialect locations in the south are further from dialect locations in the north and that there is a smooth continuum of geographic distances among locations. If linguistic

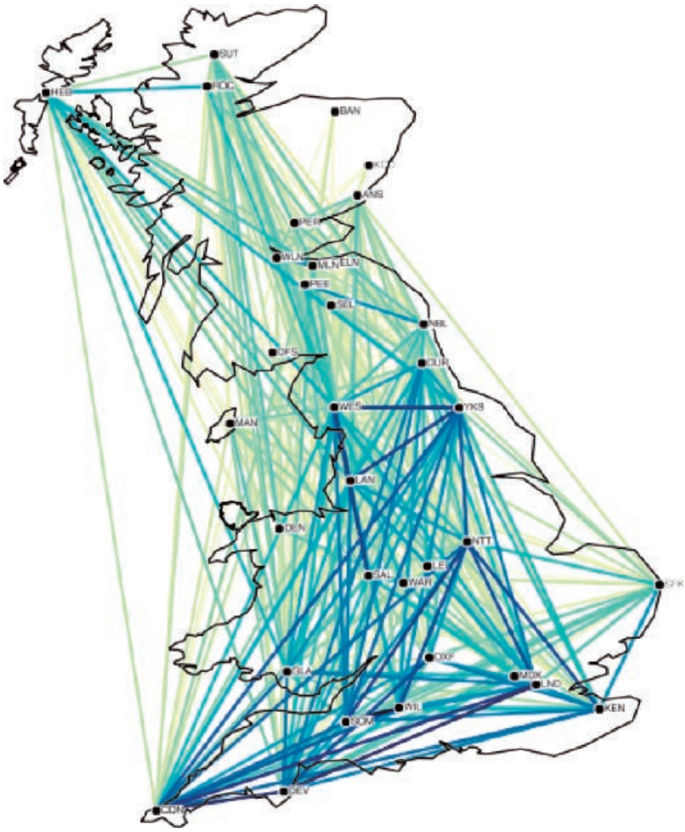


Figure 5-5 An example of the use of a geographic visualization of linguistic distance. Dialects that are linguistically closer are in darker, blueish lines and dialects that are more linguistically distant in proportionally lighter, more yellowish lines. (Szmrecsanyi 2011)

distances were similar to geographic distances we would expect the right map to be similar to the left map. But it is not, suggesting that geographic distance (as-the-crow-flies) does not seem to be correlated with linguistic distance (at least not at the macro scale).

5.2.2. Analysing Linguistic Variation Using Environmental/Geographic Variables

Linguists, cultural geographers and other social scientists are well aware that physical environments influence human activities. Toponyms, or place names, are a good example of a linguistic artefact that has the potential to provide insights into the history and culture of a place. Employing the integrative capabilities of GIS, researchers from China and the United States analyzed linguistic variation in toponyms in relation to physical features such as terrain, elevation and proximity to waterways to explore the cultural and political history of a region of southern China (Luo et al. 2000; Luo et al. 2009; Luo et al. 2018).

An example of one of their studies aimed to analyze and graphically visualize the

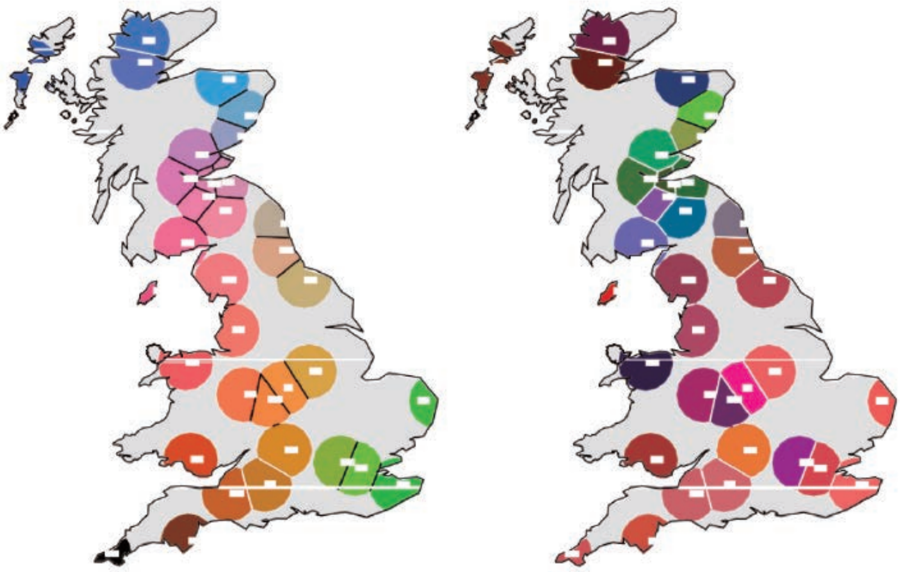


Figure 5-6 An example of cartographic visualization comparing geographical distance (left) and linguistic distance (right) based on a multi-dimensional scaling (MDS) approach for modelling distance (Szmrecsanyi 2011)

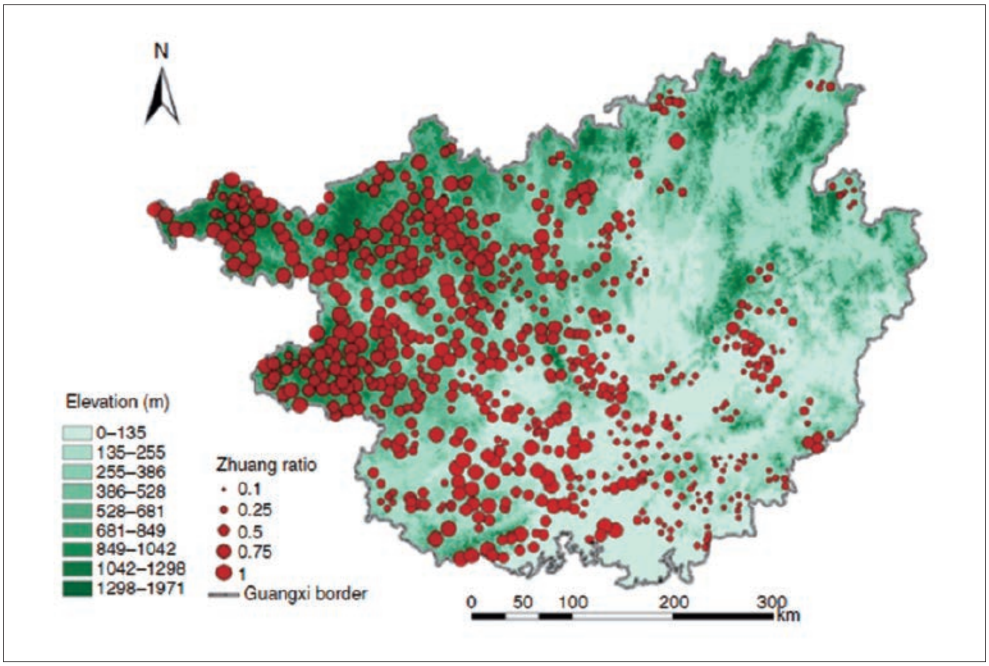


Figure 5-7 An example of the use of GIS to map geographic variation in the ratio of Zhuang to Non-Zhuang toponyms, and how the predominance of Zhuang toponyms correlates with elevation (Wang et al. 2011)

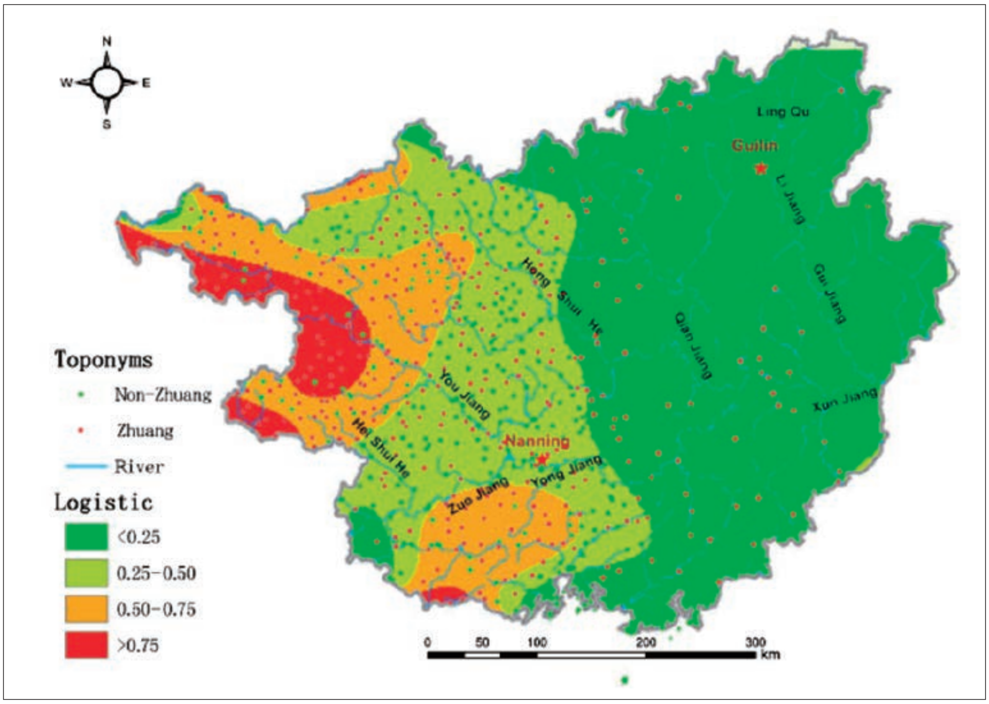


Figure 5-8 An example of the use of GIS to map geographic variation in Zhuang toponym dominance. The map shows the likelihood of encountering a Zhuang as opposed to a non-Zhuang toponym. (Wang et al. 2011)

Sinification (i.e. cultural expansion of Han Chinese) in a region originally dominated by the Zhuang ethnic group (Wang et al. 2011). Using a geocoded *Toponym Dictionary of China* they categorized 1,427 place names as either of Zhuang origin or non-Zhuang origin (i.e. Han Chinese). To quantify the dominance of toponyms of Zhuang origin they used GIS to assign a ratio of Zhuang to Non-Zhuang toponyms within a 20 km radius of each toponym location (e.g. where all surrounding toponym locations within 20 km had a Zhuang toponym, the ratio for the focal location would be 1.0) (Figure 5-7). Having converted each place location to a ratio measurement they used GIS to create a continuous surface of Zhuang toponym dominance using a common spatial interpolation method called trend surface analysis (Figure 5-8). They extended their study by evaluating the difference between Zhuang and non-Zhuang place names in relation to several environmental factors. Using GIS layers for elevation, slope, aspect, land type, distance from rivers and railways, and distance from roads and major cities, they conducted a t-test to determine whether there were significant differences between Zhuang and non-Zhuang locations for each of these variables. They found statistically significant differences for elevation, slope, land type, distance to major roads, and distance to major city. They concluded that this provided evidence that, historically, as the Han Chinese settled the region, Zhuang ethnic groups were forced to more remote, marginal lands.

5.3. The Fijian Language GIS Project

The Fijian Language GIS Project is an interdisciplinary research effort funded by The Resona Foundation for Asia and Oceania and led by Ritsuko Kikusawa of the National Museum of Ethnology, Osaka, Japan. The goals of the project are to: i) develop a GIS database of Fijian communalects collected and recorded by Paul Geraghty of The University of the South Pacific, ii) use the database to conduct research on linguistic variations from a spatial perspective focusing on horizontal and vertical evolution of language, and iii) produce information suitable for dissemination to the public through museums and other venues.

5.3.1. GIS Database of Fijian Communalects

A "communalect" is the smallest subdivision of a dialect and refers to community of native-born speakers who share a common variety of speech (Pawley and Sayaba 1971). Differences among communalects are subtle and most recognisable by native speakers as varieties of Fijian speech that indicate a person's home locality (Geraghty 1983; Pawley and Sayaba 1971). The geographic area of a communalect is small, often comprising a single village, a group of villages, or a small island (Pawley and Sayaba 1971).

Over the past 40 years Geraghty has recorded 100-word lists for over 300 communalects of Fiji. The Geraghty 100-word list is similar in concept to the *Swadesh list* used by linguists for historical and comparative language study (Swadish 1971). The list contains 100 lexical items chosen by Geraghty as a representative sample useful for comparing variation among communalects. In addition to a 100-word list for each of 300+ Fijian communalects, Geraghty has compiled a 100-word list for the Bauan dialect which has emerged since the 1840s, as the standard literary language and *lingua franca* of Fijians (Geraghty 1983; Pawley and Sayaba 1971).

To build the GIS database of communalects it was necessary to first reference communalect names to geographic locations, a process known as geocoding (Heywood et al. 2006). This was done in two ways—by assigning communalect names to a GIS dataset of villages and also by assigning communalect names to a GIS dataset of land areas known as *mataqali* (landowning units associated with Fijian clans (Crocombe 1987)).[1] As a result the GIS database is capable of representing communalects as both point and polygon features. To complete the GIS database, a table of 100-word lists for each of the 300+ communalects was imported into the database and linked to each communalect geographic feature. In total there are approximately 1200 villages in the database, each with an assigned communalect name. While there are approximately 600 communalects in total, roughly 300 communalects have 100-word lists.

5.3.2. Pilot Study: Kadavu Island

As a pilot study we used data from the island group of Kadavu to evaluate the potential for using the GIS database for a dialectometric analysis of communalect variation. Kadavu is a relatively isolated and sparsely populated island 80 km south of the main island of Viti Levu. Travel to and from Kadavu passes through two points of entry, one at an airport at

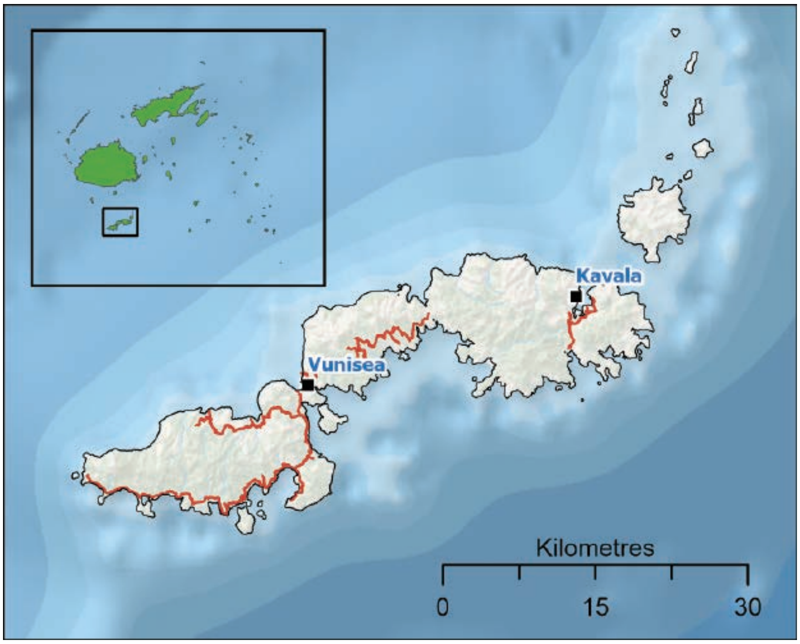


Figure 5-9 Villages of the Kadavu island group showing Vunisea and Kavala entry points (compiled by the author)

Vunisea and the other at a ferry terminal at Kavala. Automobile travel in Kadavu is limited and many people travel between villages via fiberglass boats with outboard motor. All villages are situated on the coast.

We explored the hypothesis that lexical items from the 100-word list for communities near the points of entry would be more similar to standard Fijian (Bauan dialect) than communalects further away from these points of entry. We reasoned that communities near the points of entry would be more likely to adopt artefacts of speech as a result of more frequent encounters with people moving to and from the main island. There are 59 villages, comprising 13 communalects in the Kadavu island group (Figure 5-9).

The first step was to derive a metric of linguistic distance between the lexical items for each of the 13 communalects' 100-word lists and the 100-word list for standard Fijian (Bauan dialect). This was carried out using a sequence comparison algorithm that examines the number and alignment of phonetic segments of lexical items to produce a metric of linguistic distance for each lexical item (List et al. 2018). As a result, a table was created of linguistic distances for each lexical item for each communalect, which was linked to villages in the GIS database. To visualise linguistic distance, a spatial interpolation technique called Inverse Distance Weighting (IDW) was used to create a continuous surface of linguistic distance for each lexical item (Figure 5-10).

Visual examination of four lexical times (*nikua*, *lasu*, *caka*, and *levu*) in Figure 5-10 allows us to evaluate the hypothesis that lexical items for communities near the ports of

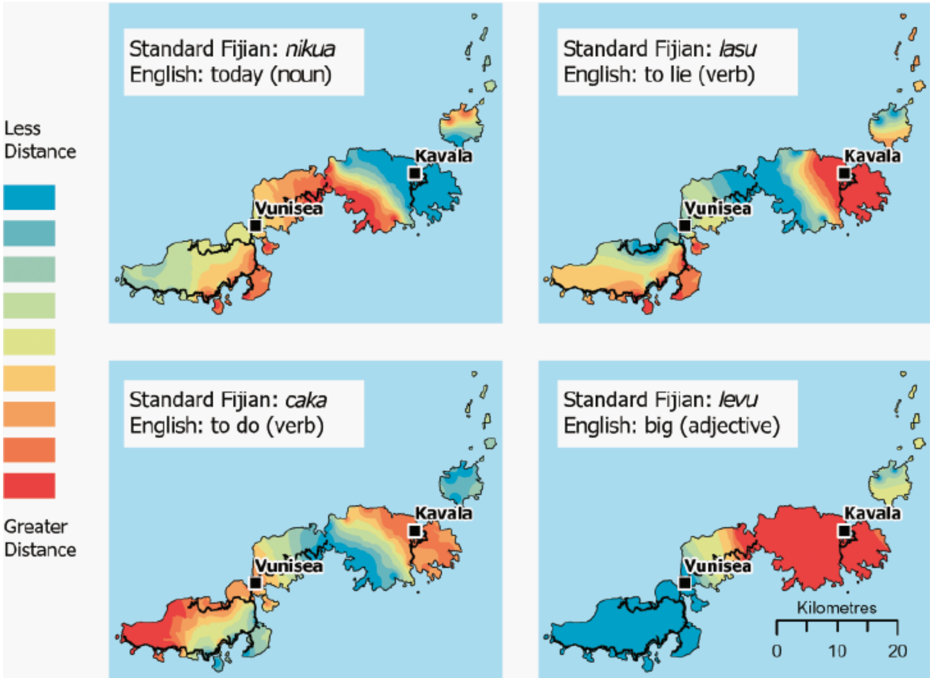
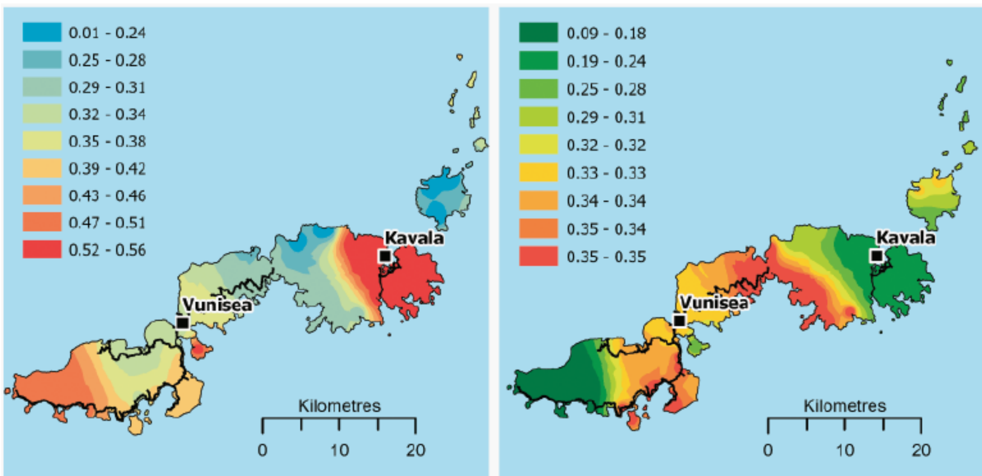


Figure 5-10 Visualization of linguistic distance to standard Fijian (Bauan) (compiled by the author)



Map A. Linguistic Distance from Standard Fijian (Ave. of 10 words)

Map B. Standard Deviation of Mean (variation in communalects)

Figure 5-11 Visualization of average linguistic distance (Map A) and standard deviation of linguistic distance (Map B) (compiled by the author) entry have greater similarity to standard Fijian than communalects more distant from the entry points. What we see is at best inconclusive, and more likely suggests the hypothesis is not supported by the data. We see, for example, that data for the lexical item *nikua* supports the hypothesis in Kavala, but not Vunisea. Likewise, the lexical item *levu* supports the hypothesis in Vunisea but not Kavala. If the data for individual lexical items are inconclusive, what if the data were visualised in aggregate? Figure 5-11 Map A presents surface based on the average linguistic distance for a selection of 10 lexical items. This can be visually interpreted to suggest that on average the communalects near Kavala are less similar to standard Fijian (i.e. greater linguistic distance) while the communalects near Vunisea tend to be neither similar nor dissimilar to standard Fijian. Figure 5-11 Map B is a visualization of the standard deviation of the mean for the same 10 lexical items. The visualization can be interpreted to suggest that communalects at the furthest ends of the island vary little in relation to standard Fijian and in the middle of the island the variation is greater.

5.4. Opportunities and Challenges

At the time of writing this chapter the Fijian Language GIS Project has been underway for over a year. Most of the 100-word lists have been entered into digital format and the GIS database of communalect names has been created, requiring one final round of quality-control. Once the database is complete several interesting possibilities in terms of mapping, information dissemination, and data analysis lie ahead. The remainder of this section briefly identifies some of these opportunities and challenges.

One of the most significant outcomes from this project is the compilation of 40 years of expert knowledge about the Fijian language(s) in a digital format that will be available to both scholars and the public. One of the questions we face is how best to make this information available. A digital atlas similar to the *Linguistic Atlas of French Polynesia* (Charpentier et al. 2015) is a logical option given the long history of linguistic atlases within academia. A major consideration, however, and an important difference between the GIS database of Fijian communalects and the Linguistic Atlas of French Polynesia, is the vast quantity of location based information in the Fiji database. Where the French Polynesia atlas mapped 19 locations (small island groups) the Fiji database has over 600 communalects spanning over 1,200 villages. In order to create an atlas that is informative and usable, we will need to consider a logical way by which to aggregate information and perhaps produce maps only for selected geographic regions of the Fiji Islands. Aggregation might be in the form of communalect groups or some other reasonable aggregate.

In addition to a digital atlas that may be primarily of interest to linguistic scholars, we must consider other mediums of dissemination that will reach the public. It is likely that these will also require some form of data reduction (i.e. aggregation) and perhaps focus on interesting aspects of the Fijian language that will be of interest to the lay person. A useful medium for reaching the public is the internet-based interactive map (Baisden et al. n.d.; Junker 2014–2019). Such an internet-based interactive map could be accessed via personal computers or from computers within a venue such as a museum. We will need to consider what aspects of this database would be most interesting to lay persons and how to make it attractive to them. Ideas might include audio and video recordings or some other form of interactive media.

Lastly the digital GIS database of Fijian communalects provides a vast resource for quantitative (and qualitative) scholarly research on the Fijian language(s). This chapter has highlighted several approaches to linguistic data analysis using dialectometric and GIS methods. These provide some ideas of what might be promising. Other approaches such as those from the field of Natural Language Processing also hold promise (Murawaki 2017). While dialectometric and other methods of data mining are designed to work with large datasets we might want to consider whether aggregating the data may be more meaningful. For example, the linguistic variation in communalects may be too subtle to detect with multivariate or other data mining techniques. It may be better to analyze the data at an aggregate level such as the communalect group. From the perspective of geographic scale we ought to consider whether it is reasonable, whether analysis at the individual island level would be more meaningful than analysing all the islands together. When considering geographic distance, is distance as-the-crow flies, reasonable? Or would it be better to measure distance along a road network or boat navigation paths? Finally it will be important to carefully consider what research questions are worth exploring and how best to address the questions of greatest value.

Note

1) GIS datasets for villages and *mataqali* were obtained from Fiji Department of Lands, Ministry of Minerals and Lands in 2012.

References

Baisden, F., C. Barker, B. Priman, M. Holden, C. Nielsen, and B. Foley
n.d. Gambay: Languages Map. http://gambay.com.au/map (Accessed January 8, 2019)

Charpentier, J., A. François, and J. Peltzer
2015 *Linguistic Atlas of French Polynesia / Atlas Linguistique de la Polynésie Française.* Berlin: De Gruyter Mouton.

Crocombe, R. G.
1987 *Land Tenure in the Pacific*, third edition. Suva: The University of the South Pacific.

Geraghty, P.
1983 *The History of the Fijian Languages* (Oceanic Linguistics Special Publication 19). Honolulu: University of Hawai‘i Press.
email: communalect definition. (July 10, 2018).

Hägerstrand, T.
1966 Aspects of the Spatial Structure of Social Communication and the Diffusion of Information. *Papers of the Regional Science Association* 16(1): 27–42.

Haynie, H. J.
2014 Geography and Spatial Anlaysis in Historical Linguistics. *Language and Linguistics*

Compass 8(8): 344–357.

Heywood, I., S. Cornelius, and S. Carver
2006 *An Introduction to Geographical Information Systems*, third edition. Essex: Pearson Education Limited.

Hoch, S. and J. J. Hayes
2010 Geolinquistics: The Incorporation of Geographic Information Systems and Science. *The Geographical Bulletin* 51(1): 23–36.

Jordan, P.
2014 Thoughts on a Concept of Language Geography. *Review of Historical Geography and Toponmastics* 9(21–22): 33–46.

Junker, M.-O.
2014–2019 The Algonquian Linguistic Atlas. https://www.atlas-ling.ca/# (Accessed January 8, 2019)

Kurath, H., M. Hansen, B. Bloch, and J. Bloch
1939–1943 *Linguistic Atlas of New England*, 3vols. Providence, RI: Brown University Press.

List, J.-M., M. Walworth, S. J. Greenhill, T. Tresoldi, and R. Forkel
2018 Sequence Comparison in Computational Historical Linguistics. *Journal of Language Evolution* 3(2): 130–144.

Luebbering, C. R.
2013 Displaying the Geography of Language: The Cartography of Language Maps. *The Linguistics Journal* 7(1): 39–67.

Luebbering, C. R., K. N. Kolivras, and S. P. Prisley
2013 Visualizing Linguistic Diversity Through Cartography and GIS. *The Professional Geographer* 65(4): 580–593. DOI: 10.1080/00330124.2013.825517

Luo, W., J. Hartmann, J. Li, and V. Sysamouth
2000 GIS Mapping and Analysis of Tai Linguistic and Settlement Patterns in Southern China. *Annals of GIS* 6(2): 129–136. DOI: 10.1080/10824000009480541

Luo, W., J. F. Hartmann, and F. Wang
2009 Terrain Characteristics and Tai Toponyms: A GIS Analysis of Muang, Chiang and Viang. *GeoJournal* 75(1): 93–104. DOI: 10.1007/s10708-009-9291-8

Luo, W., J. Hartmann, F. Wang, H. Pingwen, V. Sysamouth, J. Li, and X. Cang
2018 GIS in Comparative-historical Linguistics Research: Tai Languages. In B. Huang (ed.) *Comprehensive Geographic Information Systems,* pp. 157–180. Amsterdam: Elsevier.

Monmonier, M.
2005 Lying with Maps. *Statistical Science* 20(3): 215–222. DOI: 10.1214/088342305000000241

Murawaki, Y.
2017 Diachrony-aware Induction of Binary Latent Representations from Typological Features. *Proceedings of the 8th International Joint Conference on Natural Language Processing*, Volume 1, pp. 451–461. Taipei: Asian Federation of Natural Language Proceeding.

Nerbonne, J.
2013 How Much Does Geography Influence Language Variation? In P. Auer, M. Hilpert, A. Stukenbrock, and B. Szmrecsanyi (eds.) *Space in Language and Linguistics: Geographical, Interactional, and Congntive Perspectives,* pp. 222–236. Berlin: De Gruyter Mouton.

Nerbonne, J. and P. Kleiweg
2007 Toward a Dialectological Yardstick*. *Journal of Quantitative Linguistics* 14(2–3): 148–166. DOI: 10.1080/09296170701379260

Nerbonne, J., P. Klieweg, W. Heeringa, and F. Manni
2008 Projecting Dialect Differences to Geography: Bootstrap Clustering vs. Noisy Clustering. In C. Preisach, L. Schmidt-Thieme, H. Burkhardt, and R. Decker (eds.) *Data Analysis, Machine Learning and Applications* (Studies in the Classification, Data Analysis, and Knowledge Organization), pp. 647–654. Berlin: Springer.

Olsson, G.
1965 *Distance and Human Interaction: A Review and Bibliography*. Philadelphia, PA: Regional Science Research Institute.

Pawley, A. and T. Sayaba
1971 Fijian Dialect Divsions: Eastern and Western Fijian. *The Journal of the Polynesian Society* 80(4): 405–436.

Stoeckle, P.
2016 Horizontal and Vertical Variation in Swiss German Morphosyntax. In M.-H. Côte, R. Knooihuizen, and J. Nerbonne (eds.) *The Future of Dialects,* pp. 195–215. Berlin: Language Science Press.

Swadish, M.
1971 *The Origin and Diversification of Language*. Edited by J. Sherzer. Chicago: Aldine.

Szmrecsanyi, B.
2011 Corpus-based Dialectometry: A Methodological Sketch. *Corpora* 6(1): 45–76. DOI: 10.3366/cor.2011.0004

Wang, F., G. Wang, J. Hartmann, and W. Luo
2011 Sinification of Zhuang Place Names in Guangxi, China: A GIS-based Spatial Analysis Approach. *Transactions of the Institute of British Geographers* 37(2): 317–333. DOI: 10.1111/j.1475-5661.2011.00471.x