International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 8, pp. 2841-2853
doi: 10.25147/ijcsr.2017.001.1.191
https://stepacademic.net

Short Paper*

Language Mapping of the Cordillera Administrative Region Using Relational Model

Dalos D. Miguel
Saint Louis University, Philippines
+63 921 597 6326, ddmiguel@slu.edu.ph
(corresponding author)

John Austin T. Andres
Saint Louis University, Philippines
+63 920 527 7475, 2203262@slu.edu.ph

Heidi F. Batara
Saint Louis University, Philippines
+63 916 796 9155, 2200055@slu.edu.ph

Raquel S. Bautista
Saint Louis University, Philippines
+63 938 659 7399, 2200311@slu.edu.ph

Leonardo B. Gavino III
Saint Louis University, Philippines
+63 947 974 5828, 2200042@slu.edu.ph

Karl Adrian G. Laroya
Saint Louis University, Philippines
+63 915 343 0335, 2200591@slu.edu.ph

Rodolfo Aaron E. Mercado
Saint Louis University, Philippines
+63 915 343 0341, 2202631@slu.edu.ph

Rianne Justin R. Policarpio
Saint Louis University, Philippines
+63 999 507 0869, 2200848@slu.edu.ph

Graeven Yuri B. Videz
Saint Louis University, Philippines
+63 927 201 9980, 2201355@slu.edu.ph

Date received: January 30, 2024
Date received in revised form: March 8, 2024
Date accepted: March 17, 2024

Recommended citation:

Miguel, D. D., Andres, J. A. T., Batara, H. F., Bautista, R. S., Gavino III, L. B., Laroya, K. A. G., Mercado, R. A. E., Policarpio, R. J. R., & Videz, G. Y. B. (2024). Language Mapping of the Cordillera Administrative Region using Relational Model. *International Journal of Computing Sciences Research*, *8*, 2841-2853. https://doi.org/ 10.25147/ijcsr.2017.001.1.191

*****Special Issue on International Conference on IT Education 2023. Guest Associate Editors: **Dr. Joey M. Suba** (University of the Assumption, Pampanga) and **Dr. Bernandino P. Malang** (Bulacan State University - Bustos Campus).

Abstract

Purpose – Various studies have already done the language mapping of the different languages of the Philippines, though it only consists of the most popular languages per region. Specifically for the Cordillera Administrative Region (CAR), there is a lack of studies regarding mapping its indigenous languages to their respective communities. Resources and data about its languages are also insufficient, contributing to the language barrier problem.

Method – The study created a relational model containing CAR's languages mapped to their respective communities, along with a translation of prevalent phrases, which are used in a language mapping application. The data was collected from online sources and the Philippine Statistics Authority (PSA). Native speakers of the languages involved were also sought after to translate prevalent phrases into their respective languages.

Results – The relational model revealed that there are 22 languages spoken in CAR. The developed language mapping application displayed CAR's provinces, municipalities, languages, dialects, and translated phrases from specific languages to enable users to gain new insights about the region.

Conclusion – The study was able to create a functional prototype that can present information about the different provinces of CAR, especially its indigenous languages. Through the prototype, the study has shown the importance of preserving CAR's cultural heritage.

Recommendations – Future studies are recommended to further understand CAR's linguistic landscape. Further research about other Philippine regions' linguistic landscape is also recommended. To do so requires more data to be collected about commonly spoken languages and their dialects.

Research Implications – The created prototype may benefit tourists in knowing more about CAR and its indigenous languages by presenting information and language translations of the region. The study may also benefit future researchers and aid in preserving CAR's culture.

Keywords – natural language processing, language mapping, relational model, lexicon analysis

INTRODUCTION

The Cordillera Administrative Region (CAR) is located in the northern-central section of Luzon in the Philippines, recognized for its mountainous landscape and cultural diversity (PhilAtlas, 2020). With several indigenous languages and dialects spoken within it, charting the distribution of these languages would be beneficial for preserving and promoting CAR's cultural heritage. The process of obtaining information on the number of speakers of a certain language, the community's location, and the spoken language's features is called language mapping (Quizon, n.d.). The obtained information can be used to create language maps that provide a visual representation of the linguistics of the region. Relational Modeling is employed in this study since it focuses on the relationship between language data rather than the language data themselves (Mahavar, 2020). This makes it possible to depict intricate model interactions between communities.

Existing literature and applications on language mapping in the Philippines primarily focus on the country's most prominent languages for each region. Hence, there is a lack of study regarding mapping the communities of indigenous languages, specifically addressing the CAR. Limited resources are also recognized within the region, making it a viable tool for people, especially those having difficulties understanding words and phrases due to language barriers.

Thus, the study plans to utilize a relational model for the CAR's language map. It seeks to present the indigenous languages. For this goal, the analysis of existing results along with other findings are the primary sources of data. The study aims to create: a relational model of the CAR languages, translations of popular phrases, and a language mapping application.

The study's significance leans toward producing a comprehensive overview of CAR's language variety. The result would be valuable for tourists as they will be able to have a general understanding of CAR's languages. Additionally, this will benefit researchers who are interested in learning more about the region's linguistics and cultural legacy. It is speculated that this study would help in identifying and documenting the spatial distribution of CAR's languages.

LITERATURE REVIEW

Languages of Cordillera Administrative Region

The Philippines is considered a multilingual country due to it having more than 150 languages (Usero, 2021), as the country is culturally and historically rich, becoming home to different languages. One area in the Philippines with many heritage languages is CAR. This is located in the Northern part of Luzon Island, which covers six provinces: Abra, Apayao, Benguet, Ifugao, Kalinga, and Mountain Province.

Numerous studies have determined the reason behind the endangerment of particular languages. A study from the University of the Philippines Diliman concluded that one challenge that arose was language attrition (Usero, 2021). Also, another study from Saint Louis University discovered challenges in maintaining the heritage languages of CAR. They considered the heritage languages such as Applai, Bago, Balangao, Bontoc, Ga'dang, Ibaloi, Ibanag, Ifugao, and others, where they found out that there are three challenges: linguicism or linguistic discrimination, code-switching, and the effect of migration (Catama et al., 2016). Therefore, it is imperative to preserve these languages, as they carry the values of their ancestors and the identity of the community.

Natural Language Processing

Natural Language Processing (NLP) is a set of computational methods grounded in linguistic theories designed to analyze and interpret natural language texts. These methods operate at various levels of linguistic analysis and aim to replicate human-like language processing to perform a wide range of tasks and applications (Liddy, 2001). NLP has many applications, including retrieving information from databases based on user query relevance. Relational Database Management Systems (RDBMS) are technologies currently used for managing large volumes of information. In most cases, information retrieval from databases involves specifying search parameters by filling in a form. This can be time-consuming and tedious, especially for users unfamiliar with database query languages. NLP can make the process more intuitive and user-friendly by allowing users to enter their queries in natural language. NLP can then analyze and interpret the query and retrieve relevant information from the database without requiring the user to know complex query languages or syntax. This makes it easier for users to access and manage large volumes of information stored in databases (Samsonova et al., 2003).

This study proposes a model for collecting and sharing natural language resources associated with the languages in the CAR in the Philippines. The model is based on the Shannon and Weaver Communication Model. A working prototype of the model is presented to demonstrate its feasibility (Siquete et al., 2013).

Structured Query Language (SQL) is frequently used in RDBMS and is presently one of the most widely used languages for defining and manipulating data. Its fundamental features are utilized in all RDBMS products, with only minor variations. It is an effective tool that processes managing and querying data in relational databases (Hazboun et al., 2021).

Language Mapping in the Philippines

A few attempts have been made to create language mapping applications in the Philippines. One such is from the website where they implemented a Philippines Language Map (Translators without Borders, 2022). It is very similar to the output of this study. However, in their case, they created a map of the country and focused on the provinces. They showed the languages spoken, language percentage, and other household details from the Philippine Statistics Authority (PSA). It is a good source of data to be included in this paper. A more recent implementation was done on a website that created a language map of the Philippines to show the diversity index and the top 2 languages from each province (Thinking Machines Data Science Inc., 2010). It has limited information. However, the diversity index is a new addition to consider. The said language maps are a reasonable basis for the prototype and a good source of information as they used the census data from PSA. Nonetheless, this paper focuses on CAR and includes popular phrases from the municipalities which are more specific.

In the education field, the utilization of language mapping data for mother tongue-based multilingual education (MTB-MLE) policy was issued to aid further the creation of teaching materials (Department of Education, 2015). It is done by setting the procedures for developing, disseminating, and utilizing language mapping data. It proves the relevance and value of language mapping data in the context of education. However, the data are not widely available for specific towns and lack lexicons. Despite some studies and implementations, it still needs to be improved as most are in general analysis.

METHODOLOGY

The study utilized the quantitative approach, which explains certain phenomena by analyzing numerical data using mathematical approaches (Idowu, 2015). The phenomenon in question is the creation of the language mapping application which could be explained through a relational data model, which consists of mathematics about relations, Cartesian products, sets, and more (Johnston, 2022).

The study's target population is tourists planning to visit the provinces of CAR. CAR has been cited as one of the top tourist destinations in the Philippines (Department of Agriculture Regional Field Office - Cordillera Region, n.d.). Despite the tourism industry's focus on English, it is still advised to learn at least basic phrases of the languages of the area, which provides the need for a Language Mapping Application (Igorotage, 2017).

Research Process

The process conducted by the researchers can be seen in Figure 1. The researchers used mainly three different sources of data to conduct this study. The first data source was the Philippine Statistics Authority's (PSA) data on the languages spoken in the CAR by province and municipality. The researchers visited the PSA office in Baguio City. This provided demographic information and a list of languages and dialects spoken in each area. This was used to determine the sample size and demographics and to validate the languages used by the study population.

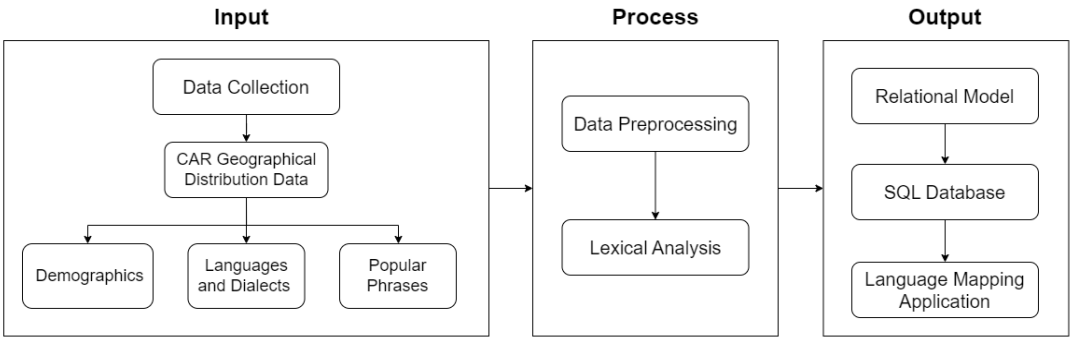

Figure 1. Research Framework

The second data source was the Translators without Borders (TWB) 2022 Philippines Language Map. This map provided information on the languages spoken in different regions of the Philippines. This data was used to identify the languages to be included in the study.

The third data source was the Thinking Machines Data Science Inc. 2010 report titled "The Language Landscape of the Philippines in 4 Maps". This provided detailed information on the language distribution and characteristics of the languages spoken in the Philippines and was used to inform the research questions and hypotheses. Other sources of information came from dependable online sources which were used to provide further information about each region in CAR and the languages and dialects used per municipality. The data gathered online for the languages and dialects were only used to supplement the data provided by the PSA office.

Furthermore, the lexicons used for the study came from the different phrases collected for each language in the CAR. These lexicons were made part of the relational model and their corresponding translation in English. The translations were mostly taken from the inputs made by native speakers of the languages involved. Other translations were taken from trusted online sources such as SIL (Summer Institute of Linguistics) Philippines.

Moreover, the relational model consists of the different provinces of CAR with their respective municipalities, languages, and common phrases. The model will take the form of an SQL database in which its records will be accessed by the Language Mapping Application.

Using the data gathered from the above-mentioned resources and techniques, the researchers developed a map application that could display the languages spoken in different regions of the Philippines, their characteristics, and their distribution. This prototype will allow the researchers to visualize the data in a more comprehensive and accessible way. Therefore, it will enable them to gain new insights into the language landscape of the Philippines.

RESULTS AND DISCUSSION

Relational Model

In developing the language mapping of the CAR, a relational model was used. Figure 2 displays the Entity Relationship (ER) Diagram of the relational database. Seven entities were created: provinces, province languages, municipalities, municipality languages, languages, dialects, and phrases. The main entity is the provinces, which other entities are related to.

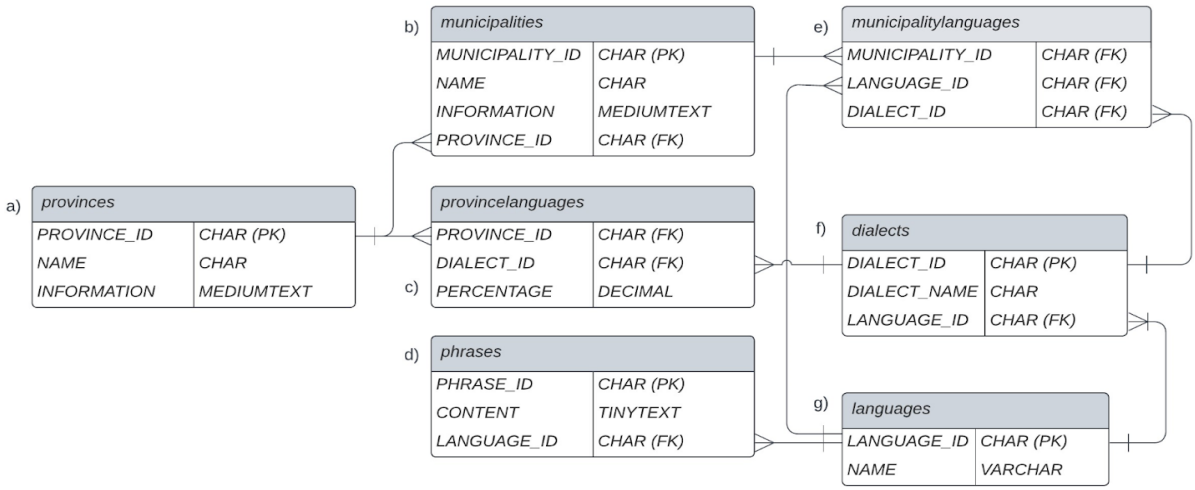

Figure 2. Entity Relationship Diagram

Each entity has a relationship with another, depending on how information is arranged in the application. It can be presented using cardinality, which indicates the numerical attribute of the relationship between the entities. The notation style used to express this cardinality is the Crow's Foot Notation. Additionally, these entities have a set of attributes in which some are considered to be primary keys which is a candidate key that will uniquely identify specific records in a relation or a foreign key that references the primary key from another table.

The first entity, provinces (see Figure 2a), has attributes such as province ID, which is considered the primary key for this relation. Also, it includes the name of the province and the information about the province. The second entity is the municipalities, in which the

municipality ID is the primary key (see Figure 2b). Similar to the attributes of the provinces table, it consists of the name of the municipality and other information. However, this relation also has a province ID that is considered to be a foreign key. The relationship between the provinces and municipalities is one-to-many, as one province can have multiple municipalities.

The third entity, languages, shown in Figure 2g, has attributes: Language ID as the primary key and the language name. Another is the dialects entity, as presented in Figure 2f, which has a dialect ID as the primary key and the dialect name. Also, it has the language ID as a foreign key that references the languages table. There is a many-to-many relationship between the municipalities and languages and between municipalities and dialects. This relationship is implemented through a junction table with a one-to-many relationship. With this, the fifth entity is the municipality languages and is considered a join table since it contains foreign keys referencing other tables (see Figure 2e). This entity has attributes: municipality ID, language ID, and dialect ID, which are all foreign keys. Therefore, a municipality can have many spoken languages or dialects, and a language or dialect can belong to many municipalities. Also, there is a one-to-many relationship between the language entity and the dialects entity, as a language can have one or more dialects.

Furthermore, the sixth entity is the province languages, which is also considered a join table since it does not contain a primary key (see Figure 2c). Instead, there are two foreign keys: province ID, referencing the province entity, and dialect ID, referencing the dialect entity. Also, it has a percentage attribute that indicates the surveyed population that speaks a particular language in a province. With that, there is a many-to-many relationship between the provinces and dialects, as a province can have many dialects, while a dialect can belong to many provinces.

The last entity would be the phrases entity, as shown in Figure 2d, which consists of the Phrase ID as the primary key, the content, which is the translated phrase based on the language, and the language ID as the foreign key that references the language entity.

Contents of the Relational Model

Given that the relationships of the entities are described, the relevant contents are as follows, gathered from PSA and other sites. CAR has a regional center, Baguio City, and six provinces which are Abra, Apayao, Benguet, Ifugao, Kalinga, and Mountain Province, having 27, 7, 13, 11, 7, and 10 municipalities, respectively.

For the languages, there are 22 prominent languages currently spoken in the CAR, namely, Applai, Atta, Bago, Balangao/Balangaw, Bontok/Finallig, Ga'dang/Baliwon, Gaddang/Cagayan, Ibaloi, Ibanag, Ifugao, Ilocano, Isneg/Isnag/Apayao, Itneg/Tingguian, Itawis, I-Wak/Iwaak/Iowak, Kankanaey, Karao, Kalinga, Kalanguya/Kallahan, Majokayong, Tagalog, English and other local dialects.

Some major languages in CAR have known dialects, such as Bontok having Central, Eastern/Kadaklan-Barlig, Northern, Southern, and Southwestern dialects. Ibaloi language has two dialects, Kabayan and Bokod dialects. Ifugao's dialects include Batad, Amganad, Henanga, Mayoyao, and Tuwali. Isneg's dialects are Isnag Calanasan and Dibagat-Kabugao. Itneg dialects are Adasen/Itneg Adasen, Banao/Itneg Banao, Belwang, Binongan/Itneg Binongan, Gobang/Gubang/Itneg Gobang, Inlaud/Illaud/Itneg Inlaud, Maeng/Itneg Maeng, Mabaka/Itneg Mabaka, Moyadan/Muyadan/Itneg Moyadan, and Masadiit/Itneg Masadiit. For Itawis, one dialect, Malaueg, was found. Kankanaey language has dialects, Mankayan-Buguias, Kapangan, Bakun-Kibungan, and Northern Kankanaey. For the Kalinga language, Balatok, Butbut, Limos, Lubuagan, Majukayang, Southern, Tanudan, Vanaw/Banao/Banao Itneg, and Mabaka Valley are the dialects. Lastly, for Kalanguya, Northern Kalanguya is one dialect found. For the other languages, there may be different dialects; however, due to limited resources, the ones listed above are the collected data so far.

The researchers considered 15 popular and common words or phrases used in CAR based on data collected through blogs and interviews, which are: "Hello", "Good Day", "Good Morning", "Good Afternoon", "Good Evening", "Welcome", "Thank you", "You're Welcome", "Please come again", "Goodbye, Sorry" or "Excuse", "Take care", "How are you?", "Yes" and "No" were translated into Tagalog, Balangao/Balangaw, Bontok, Ga'dang, Ifugao, Ilocano, Isneg/Isnag/Apayao, Itneg/Tingguian, Itawis, Kankanaey, Kalinga, and Kalanguya/Kallahan. Sample translations are shown in Table 1. The provided translations are limited due to minimal resources and translators interviewed. Nevertheless, the contents are relevant as no language mapping applications include dialects and municipalities.

Table 1. Sample Phrase Translation

Language	Phrase
English	Good Day
Tagalog	Magandang Araw
Balangao/Balangaw	Ammay wey ag-akaw
Bontok	Khawis ay akew mo
Ibaloi	Mapteng nga/ja akhew

Language Map of Cordillera Administrative Region

One of the objectives of this research is to create a language map, as presented in Figure 3. An application is created for an effective presentation of the gathered language and phrase data. The application can display three levels of depth of the map, namely the CAR-level, Province-level, and Municipality-level.

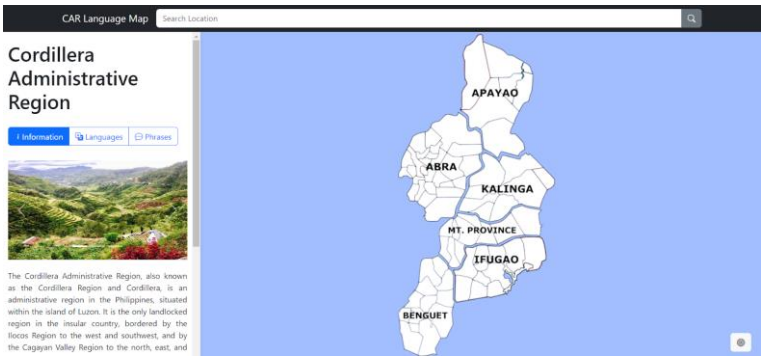

Figure 3. Cordillera Administrative Region Map

When an area is specified, it is the application's feature to display essential information about the area, the top languages spoken, and the popular phrases for each top language, as presented in Figure 4. In the Information section, pictures of the area and cultural information are provided.

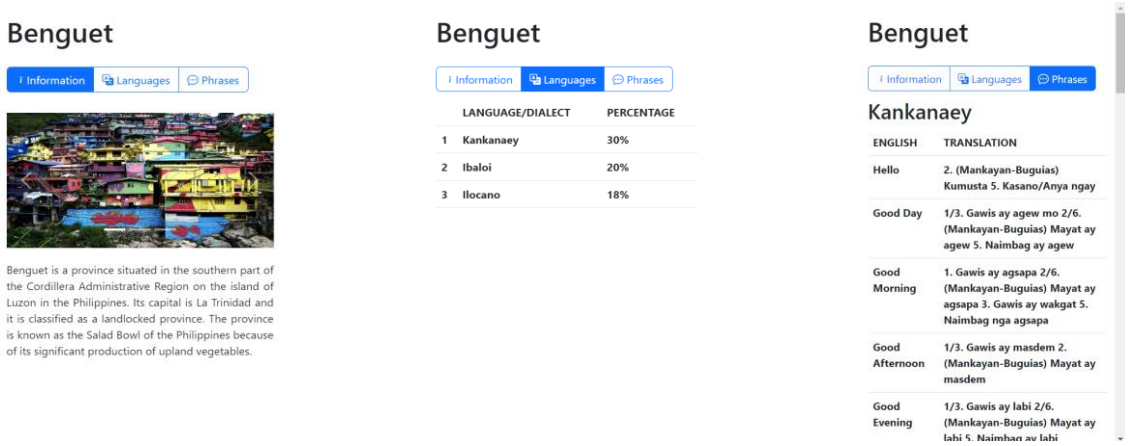

Figure 4. Cordillera Administrative Region Map Details

Next is the Language Section, which contains the top languages used in the specified area. The ranking of the languages depends on the analysis of the language data gathered. Then, the Phrases Section, contains the top phrases for each top language in the specified area. The ranking of the phrases depends on the analysis of the phrases data gathered. Paired with each phrase is their English translation Additionally, a search feature was implemented to improve the viewing experience of the users. It can search for the provinces and the municipalities (see Figure 5).

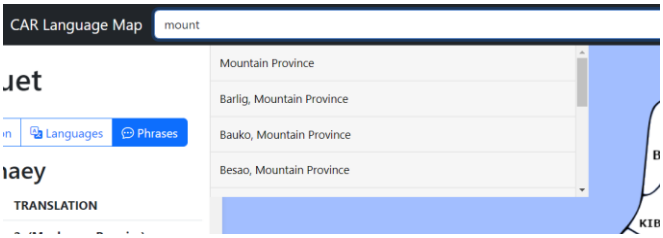

Figure 5. Application's Search Feature

CONCLUSIONS AND RECOMMENDATIONS

This study has shown the importance of preserving and promoting the cultural heritage of the CAR. In summary, the study highlights the utilization of a language map for the CAR in the Philippines and identifies the various indigenous languages that are located in it. The relational database provided seven entities namely: provinces, province languages, municipalities, municipality languages, languages, dialects, and phrases. Each of the entities was related to each other applying cardinality notation. Ultimately, the study's results are a functional prototype that generates a language map in CAR that can identify languages that are present in that certain area and translate common phrases.

Overall, this study serves as a starting point for future research to continue the language map of CAR or even other regions. In addition, it is recommended that the relational model in this study be further developed to provide a more comprehensive representation of the linguistic landscape of CAR, this can be in the form of additional entities and attributes. Finally, the study strongly suggests that more data should be collected to satisfy the requirements for translating languages, particularly on the less commonly spoken languages and dialects. Evaluation of the data and application is recommended to validate the credibility and effectiveness. Doing these allows a deeper understanding of the linguistic heritage of the country leading towards its preservation and promotion for future generations.

IMPLICATIONS

The study's functional prototype was able to visually represent CAR's linguistic data, along with other supplemental information. This function may allow tourists to know more about the region and its culture. The prototype's language map may also enhance communication between people through its translations. For the study itself, it may benefit further researchers to know more about CAR and its indigenous languages, leading to better language maps of the region and the preservation of the region's culture.

ACKNOWLEDGEMENT

The researchers would also like to thank the Philippine Statistics Authority (CAR) for allowing access to their documentation that led to the derivation of the data used in the study. In addition, gratitude is given to the anonymous translators who participated in the research. The researchers are forever grateful for the participation of these entities that led to the completion of the study.

FUNDING

The study did not receive funding from any institution.

DECLARATIONS

Conflict of Interest

All authors have declared that there is no conflict of interest.

Informed Consent

Individuals interviewed for translation consented to take part in the study and were informed that the data collected would only be used for this study. Their identity was not collected and used for this study.

Ethics Approval

The data was collected and handled accordingly based on the Philippine Data Privacy Act of 2012.

REFERENCES

Catama, B., Wacdagan, B., Tigbao, J., Ventura, G., & Viernes, M. (2016). Challenges in maintaining a heritage language. *International Journal of Research Studies in Language Learning,* 7(1). https://doi.org/10.5861/ijrsll.2017.1668

Department of Agriculture Regional Field Office - Cordillera Region. (n.d.). History. https://car.da.gov.ph/transparency-seal/history/

Department of Education. (2015). Do 55, S. 2015 – utilization of language mapping data for mother tongue-based multilingual education (MTB-Mle) program implementation. https://www.deped.gov.ph/2015/12/08/do-55-s-2015/

Hazboun, F. H., Owda, M., & Owda, A. Y. (2021). A Natural Language Interface to Relational Databases Using an Online Analytic Processing Hypercube. AI, 2(4), 720–737. https://doi.org/10.3390/ai2040043

Idowu, T. (2015). Qualitative and Quantitative research approaches. Pau-au. https://www.academia.edu/15209462/Qualitative_and_Quantitative_research_approaches

Igorotage. (2017). Cordillera Languages. https://www.igorotage.com/blog/p/9vjj9/cordillera-languages

Johnston, T. M. (2022). The Relational Model of Data: Mathematics. www.academia.edu. https://www.academia.edu/68435829/The_Relational_Model_of_Data_Mathematics

Liddy, E. D. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science*, 2nd ed. NY: Marcel Decker, Inc.

Mahavar, N. (2020). Relational Data Model and Language. https://www.goeduhub.com/4748/relational-data-model-and-language

PhilAtlas. (2020). Cordillera Administrative Region (CAR). https://www.philatlas.com/luzon/car.html

Quizon, C. (n.d.). *What are Language Maps*?. Seton Hall University. http://blogs.shu.edu/lmlc/what-are-language-maps/

Samsonova, M., Pisarev, A., & Blagov, M. (2003). Processing of natural language queries to a relational database. Bioinformatics, 19(suppl_1), i241–i249. https://doi.org/10.1093/bioinformatics/btg1033

Siquete, K. C., San Jose, T. G., & Miguel, D. (2013). A Storage and Dissemination model for Natural Language resources for Northern Luzon.

Thinking Machines Data Science Inc. (2010). The language landscape of the Philippines in 4 maps. https://stories.thinkingmachin.es/philippine-languages/

Translators without Borders. (2022). Philippines language map. https://translatorswithoutborders.org/philippines-language-map/

Usero, J. (2021). Counter-babel: Reframing linguistic practices in multilingual Philippines. Department of Linguistics - UP Diliman. https://linguistics.upd.edu.ph/news/counter-babel-reframing-linguistic-practices-in-multilingual-philippines/

Author's Biography

Dalos D. Miguel is a faculty member of the School of Accountancy, Management, Computing, and Information Studies at Saint Louis University. His specialization includes the subjects of computer Programming, Structure of Programming Languages, and Natural Language Processing. Through his expertise and knowledge, he can guide student researchers in their research endeavors.

John Austin T. Andres, Heidi F. Batara, Raquel S. Bautista, Leonardo B. Gavino III, Karl Adrian G. Laroya, Rodolfo Aaron E. Mercado, Rianne Justin R. Policarpio, and Graeven Yuri B. Videz are graduates of Bachelor of Science in Computer Science (BSCS) from Saint Louis University.