Cartography M.Sc.

Master thesis

GIS-Based Analysis and Visualisation of Indigenous-Derived Toponyms

Applied to Toponyms in Mexico

Nele Peschel

Technical University of Munich

Technische Universität Wien
Vienna University of Technology

Technische Universität Dresden

University of Twente. ITC

2022

GIS-Based Analysis and Visualisation of Indigenous-Derived Toponyms

Applied to Toponyms in Mexico

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich

Author:	Nele Peschel
Study course:	Cartography M.Sc.
Supervisor:	Dr.-Ing. Christian Murphy (TUM)
Reviewer:	Priv.-Doz. Dr.rer.nat Franz-Benjamin Mocnik (UT)
Chair of the Thesis Assessment Board:	Prof. Dr. Liqiu Meng

Date of submission: 09.09.2022

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"GIS-Based Analysis and Visualisation of Indigenous-Derived Toponyms"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Berlin, September 2022 Peschel, Nele

Acknowledgements

Thanks to my supervisor Christian Murphy who has always provided me with valuable input, critical feedback, and who has always believed in this thesis research despite ups and downs. Thanks to my reviewer Franz-Benjamin Mocnik who has always offered me his time to discuss my ideas and results.

Special thanks to cartographer and toponymist Sebastián Estremo whose work on Mexican toponymy inspired this research. Along with Sebastián, his book *Topónimos: Memoria de los pueblos en el territorio* traveled all the way from Mexico back to the Eurasian Continent where he found inspiration to create this book. It was a great pleasure to meet Sebastián in Berlin and exchange ideas about the world of cartography and toponymy.

Thanks to Evandro L.T.P. Cunha and Carlos Salvador Paredes Martínez for their interest in my research and very insightful interviews that added immense value to this research viewing the topic from different perspectives.

Thanks to Barend Köbben and Edyta Bogucka who organized the thesis semester.

Thanks to the wonderful people of the cartography programme who are all extremely talented and have made it easy to stay motivated throughout these past two years. Coming from different backgrounds and cultures they have helped me to connect the dots. Thanks to José for being my companionship on this thesis journey – literally.

And in the end, special thanks to our beautiful programme coordinator Juliane Cron who has always given her best to support everyone in this master's programme.

Abstract

Toponymy, the study of place names or geographical names, is an interdisciplinary field. While most toponymic research has been of qualitative nature focussing on a single toponym or a small set of toponyms, the development of GIS has enabled toponymy research to become more extensive (Fuchs, 2015; Tent, 2015). Applying GIS-based methods to toponymic distributions makes spatial patterns and their relationship with other variables visible. This master thesis illustrates the role and relevance of GIS in extensive toponymic research and outlines GIS-based methods commonly used in this field. It develops a toponymic classification system based on Indigenous-derived morphemes and their generic meanings and applies it to a large toponymic dataset for the study area Mexico. This classification scheme has proven useful for large toponymic datasets without etymological information and appropriate for extensive toponymy research. Furthermore, this research offers a GIS-based approach to analyse and visualise distributions of Indigenous-derived toponyms and their relationships with language distributions and the geographic environment. Applying the suggested methodology, the research has shown that selected toponyms are mainly located in areas where the language they derive from was and is spoken. Furthermore, it has shown the spatial relationship of a toponym's generic meaning with its geographic environment. Furthermore, this research suggests GIS-based methods including KDE, isopleth mapping, and hexagon mapping to compare toponyms of different feature type groups in relation to the overall toponym subset. To review and assess the results of this research, experts from the fields of linguistics and history were interviewed. This research opens possible directions for future research including intensive toponymy research to explain identified spatial patterns and relationships in the study area, the application of the suggested methodology to other study areas, and the extension of the methodology by adding other aspects to the analysis, such as multivariate approaches to improve the analysis of toponymic relationships.

Keywords: toponymy, extensive toponymy, Indigenous toponymy, GIS-based toponymy, toponym classification, spatial analysis, cartographic visualisation

Contents

List of Figures

List of Tables

Definitions

Area of Interest (AoI): Areas that exhibit an agglomeration of toponyms of a certain morpheme group within the study area. The identified AoI's are investigated in detail.

Allomorph: Morpheme variant (Lieber, 2009)

(Geographical) Feature type: Type of natural or artificial (human-made) feature in the landscape (e.g., river, mountain, or village)

Indigenous-derived: This research uses the term *Indigenous-derived* preconceived by (Blair & Tent, 2021) to refer to toponyms with an Indigenous linguistic origin

Maya: Indigenous language family spoken in Mesoamerica with large influence in toponymy of the Yucatán Peninsula (León-Portilla, 2010)

Mixtec: Indigenous language that holds a large set of toponyms influencing particularly the toponymy in Mexican state Oaxaca in Southern Mexico (León-Portilla, 2010).

Morpheme: Small segments of language that carry a meaning or are of grammatical function used to form words (Lieber, 2009)

Nahuatl: Indigenous language, also known as Mexicano according to the ethnographic map by Orozco y Berra (Biblioteca Digital Mexicana A.C., n.d.)

Purépecha: Indigenous language, also known as P'urhépecha, Tarascan (which carries pejorative connotations), or Michoacano (Paredes Martínez & Terán, 2003; Schmal, 2004)

Study area: Mexico is the study area in this research

Toponyms: Names given to geographical features or objects; also known as geographical names or place names; consist of one or more words that denote features located in a topographic space (e.g., Abdikhalikovna, 2020; Hughes, 2018; Martina, 2017; Nna, 2015; Yeginbayeva et al., 2016)

Suffix: Morpheme that attaches to another morpheme at the end (Lieber, 2009)

1. Introduction

1.1. Research Motivation and Problem Statement

Toponymy, the study of place names or geographical names, is an interdisciplinary field. It has been studied by researchers from different fields including linguists (e.g., Tent, 2017), historians (e.g., Lefebvre & Paredes Martínez, 2017), and geographers (e.g., Fuchs, 2015). These toponymists have focussed on different aspects of toponymy including etymology, the political and historical aspects of place naming practices, or migration processes that led to the denomination of a place. While most toponymic research has been of qualitative nature focussing on a single toponym or a small set of toponyms, the development of GIS has enabled toponymy research to become more extensive (Fuchs, 2015; Tent, 2015). Embedding toponyms and their spatial information into GIS brings toponyms into a spatial context where location matters. This has facilitated the involvement of cartographers in toponymy science. Applying GIS-based analysis tools and visualisation techniques to toponymic distributions makes spatial patterns and their spatial relationship with other variables visible. However, analysing toponyms quantitatively using large databases has been less explored in toponymy science than in-depth qualitative analysis. Also, the study of Indigenous-derived toponyms in an extensive manner has been a less common practice in research. Hence, there is a strong need to analyse Indigenous-derived toponyms extensively using GIS-based methods. To address this research gap, this research suggests a toponymic classification system based on Indigenous-derived morphemes and their generic meanings. It offers a GIS-based approach to analyse and visualise toponymic patterns and their relationships with language and the geographic environment.

1.2. Research Identification

The **overall objective** of this master thesis is to classify a toponymic dataset by linguistic origin, generic meaning, and geographical feature type and to apply GIS-based analysis and visualisation methods to toponymic distributions and additional datasets in order to analyse and visualise their distributions and spatial relationships with language areas and environmental variables.

The **research objectives (RO1-4)**, hypotheses (**H2-3**), and research questions (**RQ1-4**) are explained in detail in the following.

RO1:

Foremost, this research aims to develop a classification system that classifies toponymic data by linguistic origin, generic meaning, and geographical feature type. This research aims to suggest a methodology to enrich toponymic data with information about its linguistic origins, its generic meanings, and its geographical feature types which refer to the type of natural or artificial (human-made) features in the landscape, e.g., river, mountain, or village. Furthermore, the classification aims to enable GIS-based analysis and visualisation of toponyms.

> **RQ1.** *How can toponymic data be classified for GIS-based analysis and visualisation by linguistic origin, generic meaning, and geographical feature type?*

RO2:

This research aims to analyse and visualise the relationship of toponyms with historical and current language distributions by applying GIS-based methods. Furthermore, it should be assessed if and how the suggested methodology could be used to indicate a change of language use in a region. The hypothesis is that toponyms are mainly located in areas where the language they derive from was or/and is spoken (**H2**).

> **RQ2.** *How can the spatial relationship of the Indigenous language spoken in a region with the toponyms deriving from this language be analysed?*
>
> a. *Can the suggested methodology be used to indicate a change of language use in a region?*

RO3:

Furthermore, this research aims to analyse the relationship of the toponyms' generic meaning with their geographic environment. The hypothesis is that there is a spatial relationship of the toponym's generic meaning (based on morpheme) with its geographic environment (**H3**).

> **RQ3.** *How can the spatial relationship of the generic meaning of a toponym with its geographic environment be analysed?*

RO4:

Another objective is to analyse toponym distributions differentiated by their feature type group in relation to their overall count. The hypothesis is that toponyms will mainly be assigned to a feature type group their generic meaning refers to (**H4**).

RQ4. *How can distributions of toponyms of different feature type groups be analysed and compared in relation to the overall toponym subset?*

1.3. Contribution to GIS-Based Toponymy Research

This research aims to contribute to the cartography field and GIS-based toponymy research. Scientists, such as linguists or historians, may use the results to extract information on toponymic spatial patterns and to understand the toponyms' relationships with language and geographic environment in the study area. Furthermore, interviews are conducted with experts from the fields of linguistics and history to review and assess the results of this research. The results may also facilitate public understanding of toponyms, their linguistic origin, and their spatial relationships with other variables. Toponymists from different fields may apply the suggested methodology and workflows to datasets of their selected study area. This research explores and suggests different visualisation methods and applies them to different types of GIS data and to the output of spatial analysis; the cartographic visualisations generated in this research aim to be a contribution to the field of cartography.

1.4. Research Limitations

The study area is home to a variety of languages and dialects. However, this research only focusses on toponyms derived from a few Indigenous languages for which linguistic research in a toponymic context has already been conducted. Therefore, this research depends on preceding work about the study area.

Furthermore, linguistic landscapes are dynamic and prone to changes of names given to places and geographical features. However, the toponymic dataset used in this research does not include historical information about changes of toponyms; it only includes how toponyms appear today. Consequently, this research does not account for temporal changes of toponyms.

Also, the toponymic dataset does not include etymological information. Therefore, this research suggests a system that classifies toponyms into a language group based on its morphemes. However, morphemes might not necessarily be *unique* to a language; this makes the linguistic origin of toponyms ambiguous. However, this research revises third-party toponymic and linguistic literature about the study area to select Indigenous-derived morphemes. Nevertheless, the possibility of incorrect assignment in the classification remains.

Also, toponyms of Indigenous origin that are used by the Indigenous population are not guaranteed to be found in large toponymic geodatabases (Smith et al., 2012). However, the methodology could be applied to Indigenous place names extracted from, for example, participatory mapping of Indigenous place name knowledge (e.g., Smith et al., 2012).

1.5. Thesis Outline

1. Introduction: The first chapter gives a general overview of the thesis research, states the research motivation and problem, and identifies the research gap. It outlines research objectives, research questions, hypotheses, and the contribution of the research to GIS-based toponymy research and cartography. Also, it addresses research limitations.

2. Theoretical Framework: This chapter construes the framework of this thesis research. It is divided into the theoretical and methodological framework. The former gives key definitions on cartography, embeds cartography into the context of toponymy research, outlines approaches in toponymy science, presents the role of cartography and GIS in toponymy science, and introduces GIS-based research of Indigenous-derived toponyms. The latter introduces principles to classify toponymic data, touches on toponymic databases and semi-structured interviews, and outlines GIS-based methods commonly used in toponymy research.

3. Methodology: This chapter explains the research design in detail including its preparation, the development of a toponymic classification system, and GIS-based methods suggested for the spatial analysis and visualisation of Indigenous-derived toponyms. A replicable workflow is suggested.

4. Case Study: This chapter applies the suggested methodology and its workflows to the study area and describes it in detail. Prior to this, it outlines study area, datasets, software, and other research preparation steps. Furthermore, it introduces the expert interviews conducted in this research.

5. Results & Discussion: This chapter presents and discusses the results of this research. It relates the results to the research objectives, research questions, and hypotheses; embeds them into the context of GIS-based toponymy research; links them to the findings from the expert interviews; and outlines possible directions for future research. Furthermore, reliability, relevance, and limitations of the results are discussed in this chapter.

6. Conclusion: This master thesis concludes with a summary of the methodology and major findings. Furthermore, it gives an outlook for future research in GIS-based toponymy research.

2. Framework

This chapter construes the framework of this thesis research. It is divided into the theoretical and methodological framework. The former gives key definitions on cartography, embeds cartography into the context of toponymy research, outlines approaches in toponymy science, presents the role of cartography and GIS in toponymy science, and introduces GIS-based research of Indigenous-derived toponyms. The latter introduces principles to classify toponymic data, touches on toponymic databases and semi-structured interviews, and outlines GIS-based methods commonly used in toponymy research.

2.1. Theoretical Framework – Cartography in Toponymy Science

2.1.1. The Dimensions of Cartography

This research is conducted as part of the international Cartography study programme and consequently approaches the research objectives and questions from a cartographic perspective. Therefore, this section defines the term *cartography* and outlines the discipline's different perspectives.

The United Nations in 1949 defined the term cartography as follows:

> *"Cartography is considered as the science of preparing all types of maps and charts and includes every operation from original survey to final printing of maps" (United Nations, 1949, as cited in* Freitag, 1993, p. 1).

The International Cartographic Association (ICA), which considers cartography not only as the science of map production, includes technological and artistic dimensions in their definition of *cartography* as follows:

> *"The art, science and technology of making maps, together with their study as scientific documents and works of art. In this context, maps may be regarded as including all types of maps, plans, charts and sections, three-dimensional models and globes representing the earth or any celestial body at any scale" (Meynen, 1973; as cited in* Freitag, 1993, p. 1).

Georg Gartner (2021), professor for Cartography at the Vienna University of Technology and president of the ICA 2011-2015, states that the discipline of cartography relates to these different dimensions depending on the *how-questions* of cartography. If we focus on "How to communicate spatial information efficiently?" cartography relates more to science, if we deal with the question of "How to make maps?" we approach the discipline from a more technological perspective, and if we ask ourselves *"How to design* *maps?*" the focus lies more on art (Gartner, 2021). This research transfers this multi-dimensional understanding of cartography to the field of toponymy in order to address the questions shown in Figure 1.

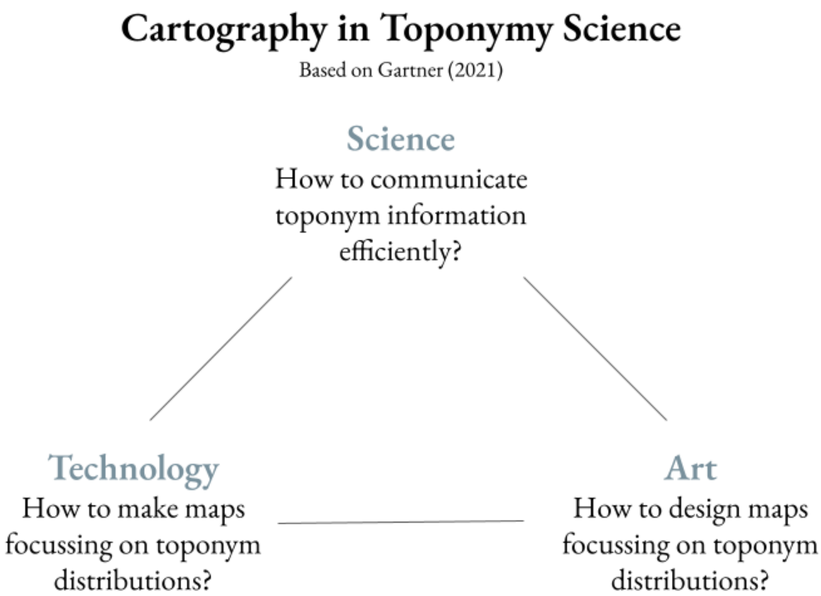

Figure 1 Cartographic questions in toponymy science formulated focussing on the three different perspectives on cartography. Based on Gartner (2021)

This thesis integrates perspectives from art, science, and technology with cartography and toponymy in the research process. The three questions in Figure 1 form the basis of the research questions (see section 1.2**Error! Reference source not found.)**

2.1.2. Toponymy Science Approaches

This section differentiates between the intensive and extensive approach to toponymy science by outlining the different characteristics of both approaches.

Toponyms, also known as geographical names or place names, are names given to geographical features or objects (e.g., Abdikhalikovna, 2020; Hughes, 2018; Martina, 2017; Nna, 2015; Yeginbayeva et al., 2016). They may consist of one or more words that denote features located in a topographic space (Abdikhalikovna, 2020). Toponyms can be – in a broader sense – divided into *habitation names* and *feature names* with the former referring to (un)inhabited localities, such as villages or towns, and the latter referring to physical features of the landscape, such as relief structures, water bodies, or vegetation features (Editors of Encyclopaedia Britannica, 2017). Toponymy often refers to a set of place names. Toponymy, as an academic field, studies and analyses the etymology, meaning, origin, and regional patterns of geographical names (Tent, 2015). The person studying toponyms is called a toponymist (e.g., Blair & Tent, 2021).

Toponymy research can be divided into large and small-scale research according to Tent (2015). Tent compares this division to the field of medicine, where *diagnostics* refers to the study of an individual case and *epidemiology* refers to the study of case patterns. Similarly, large-scale toponymy research focusses on in-depth case studies of a single toponym or a small set of toponyms whereas small-scale studies analyse regional toponymic patterns by investigating large spatial datasets (Tent, 2015). Large-scale and small-scale studies in toponymy are often considered as *qualitative* and *quantitative*, respectively (Tent, 2015). However, this distinction relates more to the type of data acquired for the toponym analysis (Blair & Tent, 2021). A small sample of toponyms can be used for an in-depth case study while a large toponymic dataset allows for quantitative analysis. However, a large dataset could also create the base for in-depth toponymic research. Therefore, for the distinction of the different research approaches the terms *intensive* and *extensive* are used (Blair & Tent, 2021; Tent, 2015). The two terms will also be used for this thesis research to distinguish between approaches that analyse toponym cases in depth *intensively* and more wide-ranging approaches that analyse spatial patterns and relationships of large groups of toponyms obtained from a dataset, gazetteer, or maps *extensively* (Figure 2).

This thesis research is extensive in nature using a large dataset to analyse spatial toponymic distributions. Examples of extensive toponymic studies include the work by Chloupek (2018); Fuchs (2015); Luo et al. (2000, 2010); Qian, Kang, & Wang (2016); Qian, Kang, & Weng (2016); F. Wang et al. (2012, 2014); and Zeini et al. (2018). Their work is outlined in detail in section 2.2.3.

According to Tent (2015) the selected approach depends on research purpose and research questions. Which approach is more important is not to be trivialised, both approaches contribute to toponymy science and to each other (Tent, 2015).

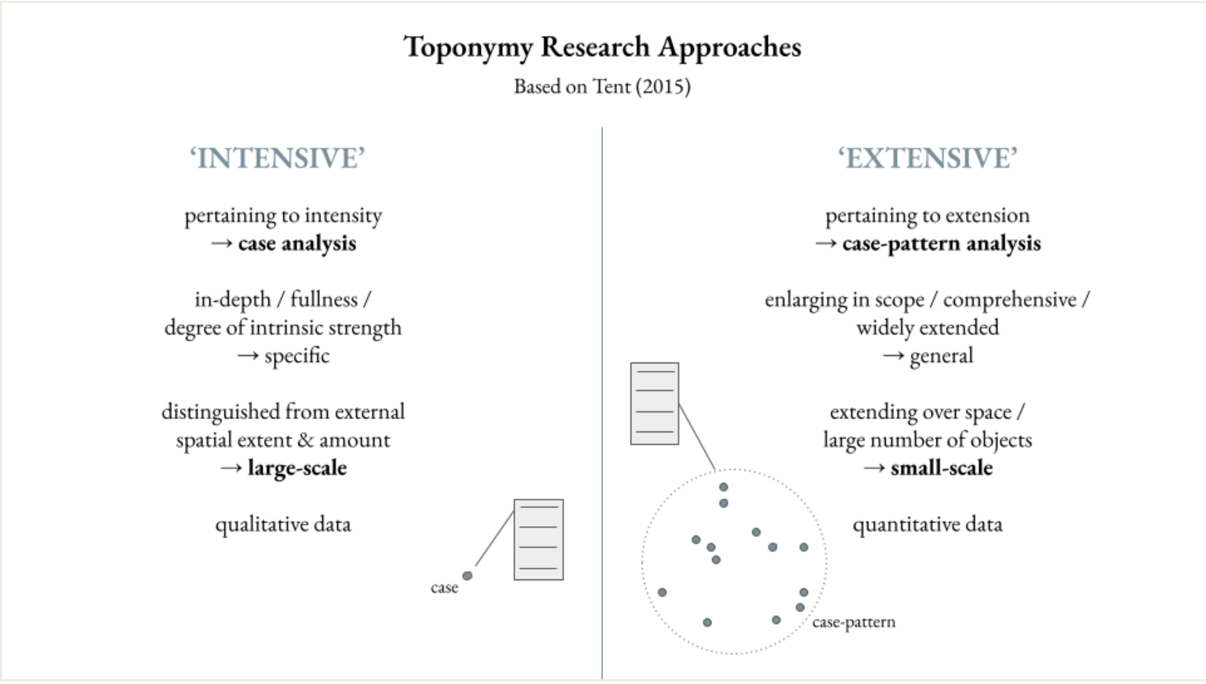

Figure 2 Toponymy Research Approaches - Intensive vs. Extensive. Based on Tent (2015)

In extensive toponymy research, toponyms act as independent variables that can be tested against dependent variables including toponym type, feature type, or environmental variables (Tent, 2015). For example, Tent (2015) presents a cross tabulation of toponyms in Australia originating from three different European languages (independent variables) and seven toponym types (as dependent variable) that reveals different place-naming practices of the European name-givers (Figure 3).

Toponym type	Percent of toponyms		
	Dutch	English	French
Descriptive	14.4%	20.2%	9.3%
Associative	12.2%	14.9%	6.3%
Occurrent	3.6%	11.3%	1.9%
Evaluative	5.0%	3.6%	1.3%
Shift	4.3%	6.0%	0.6%
Australian Indigenous	0.7%	0.4%	0
Eponymous	59.7%	43.5%	80.6%

Figure 3 Place-naming practices of European powers in Australia 1606-1803 (Tent, 2015, p. 71 Table 2)

Research in toponymy has traditionally been conducted by linguists and historians with a complex understanding of language and place naming practices (Tichelaar, 2002). While linguists are interested in studying the taxonomy of languages enclosed in toponyms, historians might want to understand the political and historical aspects of place naming and find out about migration or cultural exchange patterns through toponymic research (Tichelaar, 2002). Therefore, the aim has been to gain an *intensive* understanding of the etymology, meaning, and origin of a single toponym or a small set of toponyms by investigating case studies on a micro-level (Tent, 2015). Papers on toponym analysis taking the extensive approach are presented less often at toponymic conferences (Tent, 2015). This is supported by a survey indicating that more than half of all publications on toponymy from 1952 to 2014 published in the journal Names (54.4%) have been categorized as qualitative, while only 15.6% fall into the category of quantitative toponymy research (Tent, 2015). 7.1% are a mixture of both, qualitative and quantitative research, and 22.9% refer to theoretical papers on toponyms. However, Tent (2015) states that extensive toponymy science has been facilitated thanks to modern resources, such as geographical name databases, online gazetteers storing large amounts of toponymic data, or computer programs, such as Excel or SPSS.

2.1.3. Cartography and GIS in Toponymy Science

This section outlines the role of cartographers in toponymy science and explains how the development of GIS has changed this role, allowing for extensive toponymy research.

Traditionally, cartographers have had different roles in toponymy. As toponym creators, they were sometimes directly involved in the naming process of geographical features or whole land masses. For example, Martin Waldseemüller introduced the choronym *America* to the world map in 1507 (Włoskowicz, 2020). As toponym collectors who receive geographical name data from linguists, geographers, or surveyors, cartographers have had indirect responsibility for choosing the best source in order to correctly represent geographical names (Czarnota, 1930; as cited in Włoskowicz, 2020).

However, cartographers' involvement in toponymy science has changed with the development of GIS. GIS is a spatial system that creates geographic data; stores and manages it in a database; analyses spatial relationships, patterns, and trends; and produces visualisations of geospatial information in the form of maps which are *"the most persuasive GIS output"* (GISGeography, 2022; O'Sullivan & Unwin, 2010, p.56). Esri, the international supplier of GIS software, describes GIS as a system that connects geospatial data with other types of information to perform spatial analysis with the aim of identifying patterns and relationships of geographic nature (Esri, n.d.). Also Dent et al. (2009) state that besides visualisation and map production, geographic information system's real strengths are spatial data management and spatial analysis. Digital GIS-based cartography has enabled cartographers and map makers to visualise already collected and codified toponyms stored in databases along with its spatial information taking little effort compared to the times before digital map making (Włoskowicz, 2020). Furthermore, it has created the possibility for extensive toponymy approaches (Fuchs, 2015). We know that by embedding toponyms and their location into a GIS, cartographers bring toponyms into a spatial context where location and their spatial relationship with other variables matters. Applying cartographic methods to toponymy – what Tichelaar (2002) defines as part of "applied cartography" – has the aim to analyse and visualise toponymic distributions. GIS tools allow to analyse these spatial toponymic patterns and their relationship with the environment which presents a huge progress in toponymy science (Zhang, 2012; Chen et al., 2014; as cited in Zeini et al., 2018). Consequently, we can state that digital GIS-based cartography enabled the cartographer to become a toponym analyst or explorer.

The insights obtained from GIS-based extensive toponymic research are manifold as they offer us knowledge of regional patterns and relationships of certain toponymy types (Chloupek, 2018; Cooper, 2020; Tent, 2017), on the origin and development of ethnic groups (Qian, Kang, & Weng, 2016), settlement patterns (Zeini et al., 2018) and regional geomorphology or topography (Hughes, 2018; Luo et al., 2010).

2.1.4. GIS-based Research of Indigenous-Derived Toponyms

This section outlines the necessity of GIS-based research of Indigenous-derived toponyms and gives reasons for the lack of extensive toponymy research in this field.

Wang et al. (2019) consider geographical features as valuable components of regional culture as the region's history and the language of the toponym's name-giver is stored in its denomination. Because

Indigenous people and their languages form an active part in the constitution of a region's toponymy, toponyms of Indigenous origin are valuable assets to retrace the history of a geographic space (Israel et al., 2016). According to Zhu et al. (2018, p. 2), Indigenous toponyms are *"living heritages"* storing language and culture in their names. Furthermore, current research topics that concern the empowerment of Indigenous people, decolonization processes, and land claims strongly recognize the importance of studying toponyms of Indigenous origin (e.g., Cole & Hart, 2021; Nna, 2015; Palmer & Korson, 2020). With Indigenous languages becoming endangered or even extinct, toponyms turn into cultural and historical remnants which makes studying geographical names given by the Indigenous population part of preserving regional culture (Israel et al., 2016).

Nevertheless, there is still few extensive toponymic research focussing on the toponym's linguistic origin. One reason for this is certainly the lack of etymologic information in large toponymic databases due to the difficulty to identify the etymological origin of geographical names (Yeginbayeva et al., 2016). Another reason, particularly referring to toponyms of Indigenous origin, could be that geographical names as they are utilized by the Indigenous population are not stored in large databases or the information stored in the database differs from the Indigenous denomination (Smith et al., 2012). Amongst other reasons, Indigenous toponyms have therefore rather been analysed intensively using data obtained from field work or historical maps (e.g., Martina, 2017; Smith et al., 2012). Data from field work assures that the geographical names are utilized by the local Indigenous population (Hughes, 2018). Furthermore, obtaining geographical denominations by field work has participatory character as it includes the local Indigenous population directly in the data collection process. However, there is some examples on extensive toponymic analysis focussing on Indigenous toponyms: Tent (2017) used national gazetteer information for the study of Indigenous and Introduced (non-Indigenous) toponyms in Australia, New Zealand, and Fiji. Wang et al. (2012) used the data from the toponym dictionary series of China (Cui, 1999).

2.2. Methodological Framework – Toponym Classification and GIS-based Methods in Toponymy Science

2.2.1. Toponymic Classification

To retrieve any practical value, extensive toponym studies must be based on toponymic data that is structured in a systematic and functional way (Blair & Tent, 2021). Developing and creating a so-called *toponymic typology* includes the selection of toponym categories that are comprehensive and effective with regards to the research purpose (Blair & Tent, 2021). This subchapter presents a selection of principles according to which toponyms can be classified.

2.2.1.1. Natural Versus Artificial Geographical Features

This section deals with the dualistic notion of toponyms as denominations of either natural or artificial geographical features.

Considering various perspectives, Abdikhalikovna (2020) studied different principles in toponymy classification in detail. According to the researcher's findings, one way to classify toponyms is to divide them into two main groups: (1) *toponyms of man-made objects* and (2) *toponyms of natural objects*. Toponyms of manmade objects can be subdivided into, e.g., *horonyms* (territories of definite borders), *oykonyms* (names of settlements), and many other subgroups. Toponyms belonging to the group of natural objects can be further subclassified into, e.g., *hydronyms* (water objects including rivers, lakes, swamps, etc.), *oronyms* (relief elements including mountain, hills, hollows, etc.), or *drymonyms* (forests) (Abdikhalikovna, 2020). Similar is the broad division by Editors of Encyclopaedia Britannica (2017) given in 2.1.2 dividing toponyms into (1) *habitation names* and (2) *feature names*. Tent (2017) also differentiates between names given to *natural* features and *non-natural* features. Within these two broad main groups toponyms can be further categorized by themes (topographical feature theme) shown in the cross tabulation (Figure 4) that includes the percentage of toponyms that are classified as *Indigenous* or *Introduced* (Tent, 2017).

		Topographical feature theme	Percentage	
Country	Natural vs non-natural		Indigenous	Introduced
Australia	Natural	1 Marine	12.1	87.9
		2 Inland water	23.7	76.3
		3 Relief	20.2	79.8
		4 Vegetation & Desert	27.2	72.8
		Total percentage	21.1	78.9
	Non-natural	5 Constructed (human artifice on the topography)	30.0	70.0
		6 Civic (administrative units)	44.6	55.4
		Total percentage	33.6	66.4

Figure 4 Distribution of natural vs. non-natural feature toponyms (Tent, 2017, p. 207 Table 2 [edited])

In conclusion, a dualistic approach can differentiate toponyms broadly by denominations including artificial and human-made geographical features including localities; and natural geographical features including rivers, mountains, hollows, or forests. However, subgroups to differentiate geographical features within each of these two classes often follow this broad division.

2.2.1.2. Morphemes as Toponym Classifier

This section deals with morphemes as indicator of linguistic origin and as a tool to classify toponyms. The definitions of morphemes and the different types of morphemes relevant for this thesis research are based on the book *Introducing Morphology* by professor of linguistics Rochelle Lieber (2009).

Morphemes are small segments of language that carry a meaning or are of grammatical function used to form words (Lieber, 2009) They are divided into two broad classes, *free* morphemes that can stand alone and *bound* morphemes that are typically attached to other morphemes, such as *affixes*; *affixes* can be *prefixes* or *suffixes* that attach to another morpheme either at the beginning or at the end, respectively (see Figure 5). Morphemes to which *affixes* are attached to are called *base* (or *stem*) morphemes. *Base* morphemes create the core of a word; they are *bound* when attached to other morphemes. However, they also occur as *free* morphemes.

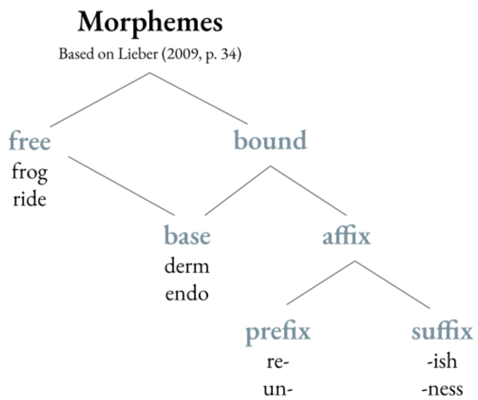

Figure 5 Types of morphemes (Simplified). Based on Lieber (2009, p.34 Figure 3.1 [edited])

Because morphemes construct and structure language, they are indicators of the linguistic origin of words. Morphemes as part of geographical feature names carry a meaning often relating to the geographic environment of the toponym's location as shown and described by Zeini et al. (2018) and Yeginbayeva et al. (2016). An important work that highly influenced and inspired this master thesis is *Toponyms are the memory of the territories' people* (translated from Spanish: *Los topónimos son la memoria de los pueblos en el territorio*) performed by Mexican cartographer Sebastián Estremo who created a series of static maps of Mexico. The dot maps visualise localities containing one or more morpheme(s) or letter(s), show its/their count, and give textual information on its/their linguistic origin, history, and meaning (Figure 6) (Estremo et al., n.d.).

Figure 6 Mexican toponyms including the suffixes -cuaro, -taro, -huén, and -uato visualised as dots (Estremo et al., n.d.).

A similar yet different project is *The small atlas of settlement names in Germany* (translated from German: *Kleiner Atlas der Siedlungsnamen Deutschlands*) published by the Leibniz Institute for Regional Geography that filters and visualises toponyms based on morphemes to give insights into the history of German place names (Leibniz-Institut für Länderkunde, 2021). The project states that toponyms are an indicator of the natural environment and the time of the foundation of a locality. However, it is not always possible to track down the origins of ancient place names due to the absence of official documents and name alternations during the past (Leibniz-Institut für Länderkunde, 2021). Tent (2015) also confirms that information on etymology and meaning of ancient toponyms in Europe is often lacking. The web application filters toponyms by morphemes that can be selected by the user, visualises their distribution in the form of a dot map, compares them by using distinct colors, and informs about the settlement history. The example in Figure 7 filters the settlement names by the suffixes -rod (red dots) and -rode (blue dots) using the web application.

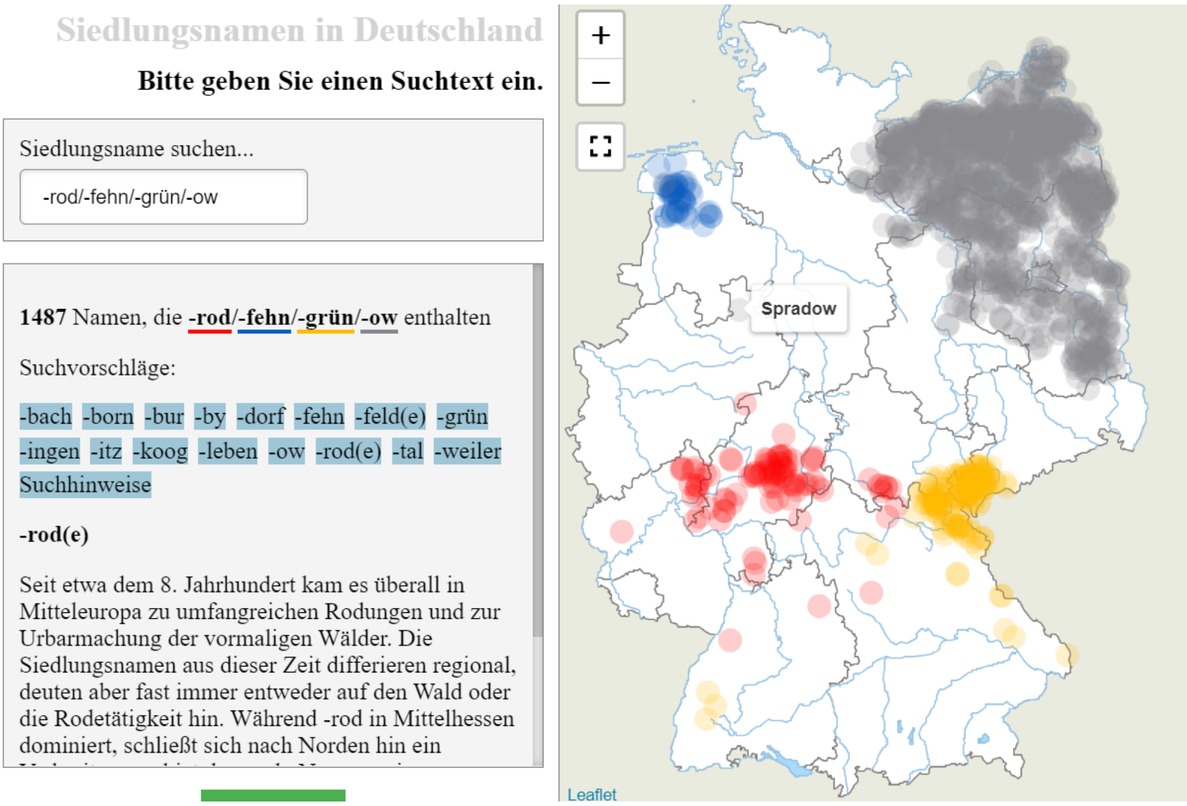

Figure 7 Section of the web application 'The small atlas of settlement names in Germany' (Leibniz-Institut für Länderkunde, 2021).

Both projects, the toponymic maps by Estremo and the atlas of settlements, offer interactive or static dot maps that feature inhabited places containing selected morphemes as point features. They engage the user to explore the linguistic origins and history of toponyms by setting the toponym groups into context, and to get to know about the people who inhabited the regions.

The research conducted by Luo et al. (2010) analyses the spatial distribution of toponyms including the three words in Tai language, *Muang* which means flat *basin*, *Chiang* which means *town*, and *Viang* which protects and defends a *Chiang*, and compares it to terrain characteristics. The results show that the distribution of these toponyms reflects the geographic environment that is associated with the meaning of the toponym denominations. Furthermore, it demonstrates the existence of a common place-naming history influenced by their cultural practices, such as wet rice agriculture.

The projects and scientific work conducted by the Leibniz Institute for Regional Geography (Leibniz-Institut für Länderkunde, 2021), Sebastián Estremo (Estremo et al., n.d.), and Luo et al. (2010) are good examples for extensive toponymy research using morphemes to create toponymic classes.

2.2.1.3. Classification Based on Generic Meaning

This section shows that the generic meaning of toponyms or of its morphemes can be useful for toponym classification as the meaning often indicates the type of geographical feature the toponym denominates.

Qian, Kang, & Weng (2016) classify toponyms according to their generic names that are unique to an ethnic group and that carry a meaning. This way, toponyms are assigned to an ethnic group and their language or dialect, and to a feature type group according to their meaning. Since the generic or descriptive names are mostly occurring as morphemes, this classification approach is similar to the approach described in the section beforehand (see 2.2.1.2 Morphemes as Toponym Classifier). Zeini et al. (2018) use a combined complementary approach to classify toponyms for their quantitative GIS-based spatial analysis. The study includes a classification approach based on geographical character, the so-called *geographical-based approach*. This approach consists of three stages where classes are defined based on meaning of generic terms, geographical location, and linguistic analysis. Using this approach shows that for most toponyms their generic meaning reflects the geographical environment (Zeini et al., 2018). Also, Yeginbayeva et al. (2016, p. 12) describe that *"the names of each land and water are like documents confirming the nature of this region"*. Luo et al. (2010) also state that the generic meaning of Tai toponyms often reflects their geographic environment. Another research states that toponyms in the United States usually do not relate to the geographical feature they represent (Chloupek, 2018). However, toponyms belonging to the category of Indigenous toponyms in Nebraska have shown the opposite (Chloupek, 2018).

To sum up, the generic meaning of toponyms or their morphemes can be an indicator for the geographic environment the toponyms are located and can be used to classify toponyms.

2.2.1.4. Classification Based on Indigenous Origin

This section introduces toponym classifications based on their Indigenous origin.

In toponymy research focussing on toponyms of Indigenous origin, one principle is to broadly classify toponyms based on their linguistic origin, like in the case study of Abdikhalikovna (2020) where toponyms are grouped into the classes *native*, *borrowed*, or *hybrids*. Wang et al. (2014) group the multiethnic toponyms in the Yunnan region in southwestern China into three language branches for which each group contains toponyms originating from multiple minority languages. The groups *Zang-Mian* and *Zhuang-Dong* include languages of native ethnic minority groups in the region, whereas the third group includes *Han* (Chinese majority) toponyms. Many studies take a binary approach and divide toponyms into two groups: *Indigenous*; and *non-Indigenous* or *introduced* toponyms Tent, 2017; Wang et al., 2012). However, in a revised typology Blair & Tent (2021) state that a more proper term for toponyms originating from Indigenous languages would be *Indigenous-derived*. This is because toponyms of Indigenous origin could be new toponyms introduced into a national's toponymy because they existed as Indigenous names in the past – termed as *Copied*; or they are new toponym inventions based on Indigenous words that are supposedly euphonious or semantically appropriate – termed as *Innovative* (Blair & Tent, 2021). Interviewee Evandro L.T.P. Cunha states that in the case of Brazil there is many toponyms – given to towns, neighborhoods, or even bigger cities – that were named after Tupi language, a language spoken by the Native people in South and Southeast Brazil. However, Cunha says that these places were not named by Tupi people themselves, but their names were adopted from Indigenous language and given to places by Europeans who found value in the country during the romantic era in the 19th century (Cunha, 2022, see transcription of expert interview in section 7.1.1). Therefore, *Indigenous-derived* might be a more proper term, because labelling a toponym as *Indigenous* might suggest that it was named directly by the Indigenous population. Also, Evandro L.T.P. Cunha prefers calling these toponyms *toponyms based on Indigenous languages* in Portuguese (translated from Portuguese: *"baseada em línguas indígenas"*) (Cunha, 2022, see section 7.1.1). Additional information could help to differentiate between toponyms named by Indigenous people and those not named by Indigenous people (Cunha, 2022, see section 7.1.1). However, particularly in extensive toponymy research the author considers the integration of area-wide and comprehensive knowledge of historical naming practices to proof a toponym's linguistic origin rather difficult.

Because of the uncertainty of whether a toponym carrying an Indigenous-based morpheme was given by Indigenous people, this research uses the term *Indigenous-derived toponyms* preconceived by Blair & Tent (2021). As this research does not hold information on who named places, the term *Indigenous-derived* aims to imply that the toponyms might not have been named by the Indigenous people themselves. In this research, the term refers to toponyms that carry a morpheme predominant in the toponymy of an Indigenous language.

2.2.2. Toponymic Databases

This section covers a small selection of databases used in toponymy research.

The use of toponymic databases is particularly interesting for extensive research. For the spatial analysis of German and German-related toponyms in the Midwest region of the United States, Fuchs (2015) used toponymic information from the Geographic Names Information System (GNIS) developed by the US Board on Geographic Names (BGN) and the US Geological Survey (USGS). Similar to the GNIS database, the GEOnet Names Server (GNS) provides access to the database by the US National Geospatial-Intelligence Agency (NGA) and BGN for geographical feature names outside of the United States. Luo et al. (2010) used data from the GNS database to investigate relationships of Tai toponyms with terrain characteristics. Another important source for toponyms in China is the toponym dictionary series of China (Cui, 1999). The database consists of information from local gazetteers compiled and updated by provincial place name offices. This compilation has highly supported and facilitated toponymy research in China (Wang et al., 2014). The dictionaries explain the linguistic origin of the toponyms and in some cases include information on the era of their first appearance (Wang et al., 2014).

2.2.3. Overview of GIS-based Toponymy Research

This section gives a chronological overview of selected extensive toponymy research using GIS-based analysis and visualisation methods. Some of the introduced methods and others will be discussed in detail in 2.2.4. Most of the toponymic research presented in this section applies GIS-based methods on toponyms originating from minority ethnic groups.

Toponymy research focussing on toponyms originating from minority languages has particularly been conducted by Chinese scholars in Tai-speaking southwestern China; Luo et al. (2010) also include regions in Vietnam, Laos, Burma, and Thailand. A GIS approach to investigate spatial patterns of toponyms based on their linguistic origin is dating back to 2000, when Luo et al. (2000) examined toponyms related to wet-rice farming of Tai minority groups. They suggested Guangxi-Guizhou in southern China to be the region of origin of Proto-Tai, the ancestor of Tai languages. The region has primarily been settled by the Zhuang people, the largest minority group in China. The study concludes that GIS-based analysis and visualisations have big potential in linguistic geography research as they support the explanation of ethnic group distributions. Further research in Tai toponymy investigated the relationship of Tai place names with terrain characteristics using GIS-based techniques (Luo et al., 2010). The hypothesis that Tai toponyms are mostly located in areas with a particular topography suitable for their cultural and economic practices, could be corroborated by identifying significant spatial patterns of Tai toponyms using kernel density estimation (KDE), the compound topographic index (CTI), statistical t-tests of terrain parameters in regions of high and low toponym concentrations, and the calculation of the nearest distance to rivers (Luo et al., 2010). The KDE result visualises Tai toponym concentrations as a ratio among all place names as shown in Figure 8. The map visualises the relative concentration of Muang toponyms; in Tai, *Muang* means flat basin (Luo et al., 2010).

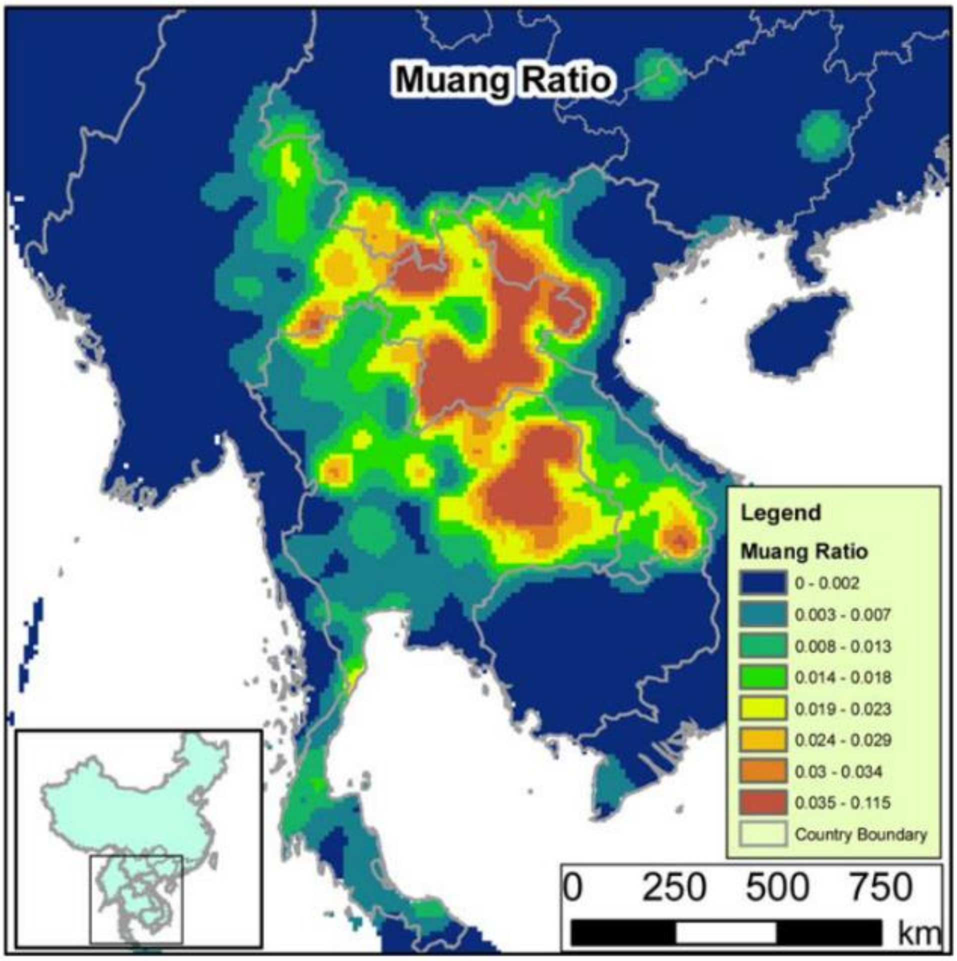

Figure 8 KDE result visualised as an interpolated surface shows the relative density of Muang toponyms (Luo et al., 2010, p.8 Fig.5)

The study of Wang et al. (2012) developed a GIS database of place names at different scales to analyse spatial distributions of *Zhuang* and *non-Zhuang* toponyms in Guangxi. Furthermore, regression analysis examined the dependence of Zhuang concentration patterns with various explanatory variables of environmental and human nature, such as town locations, road networks, and elevation (Wang et al., 2012). Furthermore, a choropleth map was used to visualise Zhuang population in this research. To represent spatial trends, Wang et al. (2012) used spatial smoothing whose result is visualised by a proportional symbol map representing toponym ratios (Figure 9). Also, trend surface analysis – a spatial interpolation method – was used to visualise the probability of toponyms originating from a minority language as a continuous probability surface using the isopleth mapping method (Figure 10) (Wang et al., 2012). In addition, circular dots that differentiate between Zhuang and non-Zhuang toponyms by different colors are placed on top of the isopleth map.

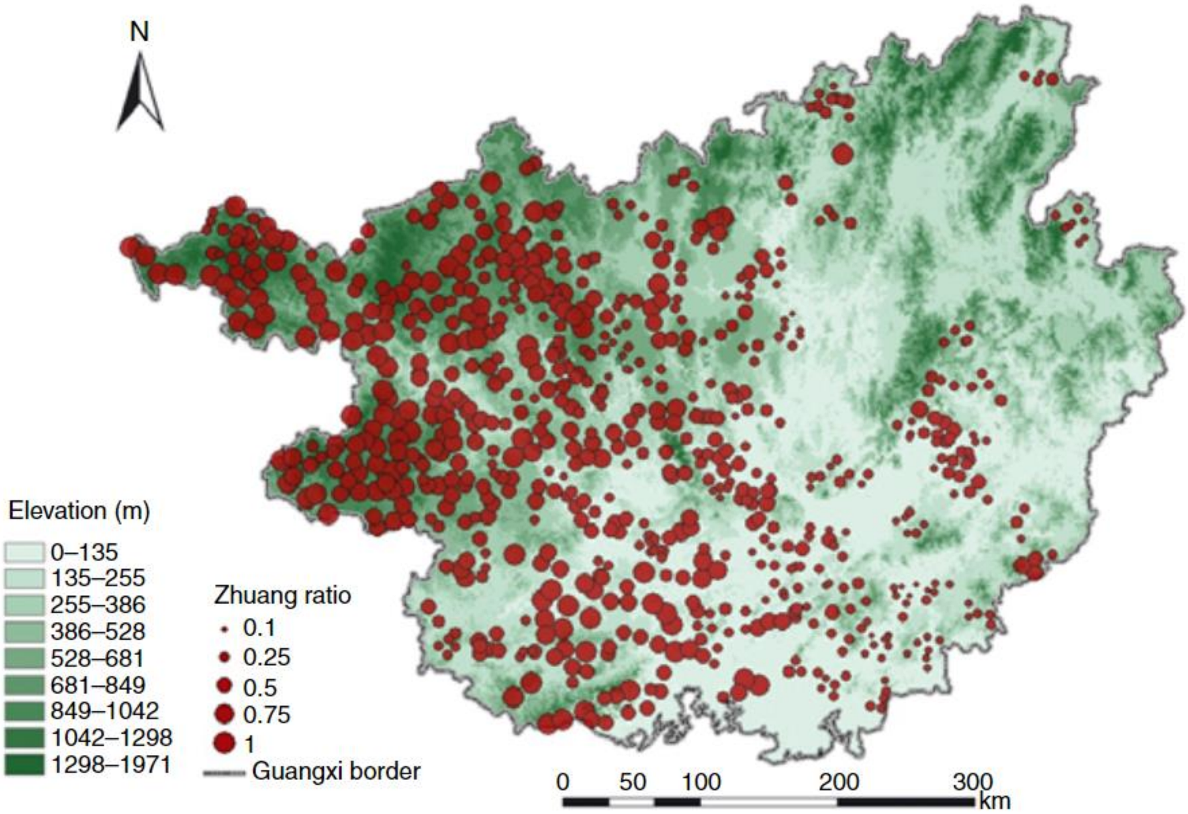

Figure 9 Proportional symbol map visualises toponym ratios (Wang et al., 2012, p.8 Figure 5)

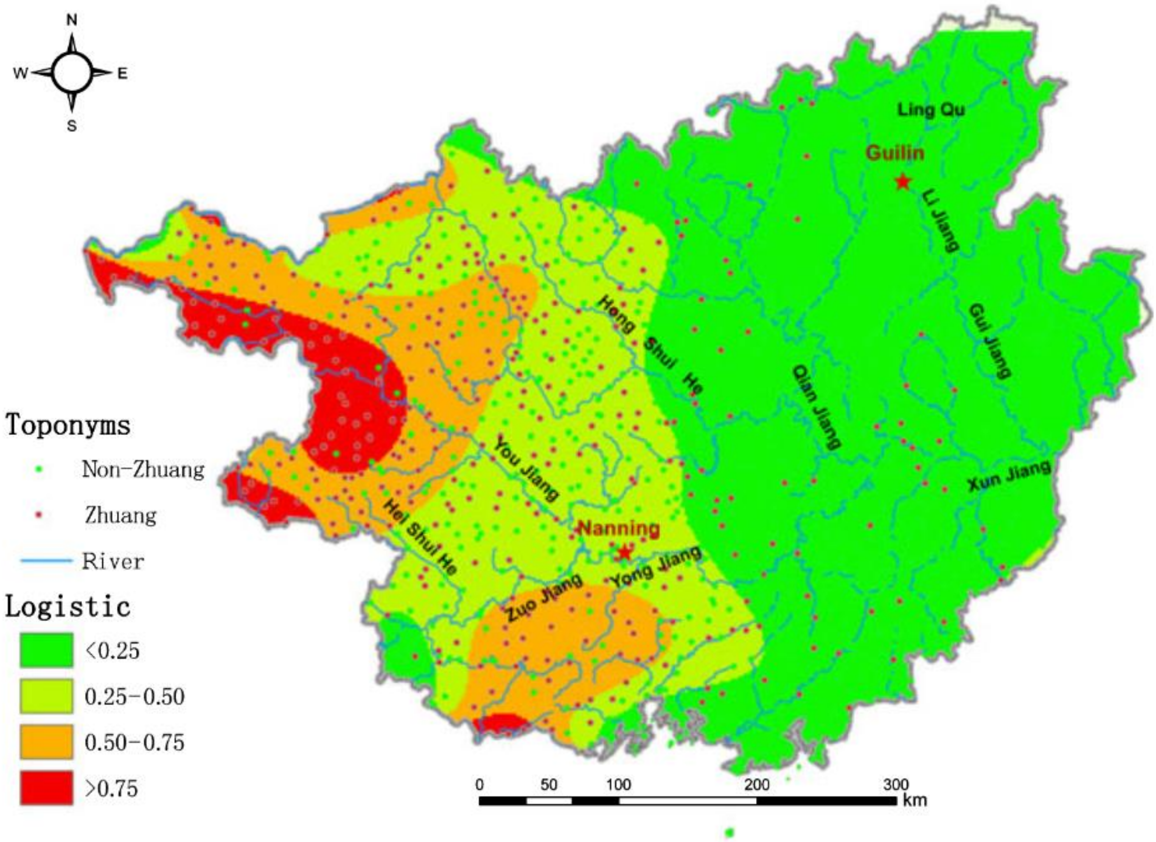

Figure 10 Trend surface analysis visualises the probability of toponyms originating from a minority language as isopleths (Wang et al., 2012, p.8 Figure 6)

In contrast to the previously mentioned scientific work, Wang et al. (2014) studied toponyms originating from multiple ethnic language groups in Yunnan, another region in southwestern China inhabited by various ethnic minority groups. Wang et al. (2014) grouped place names based on their multiple linguistic origins (the information was exported from a toponymic dictionary and stored in a geodatabase). Wang et al. (2014) used IDW (Inverse Distance Weighting) as spatial interpolation method to visualise the variation of relative concentrations of two minority toponym groups compared to Han (Chinese majority) toponyms (Figure 11 and Figure 12). The maps show the interpolation results as continuous surfaces ranging from blue (low relative concentration of minority toponyms) to red (high relative concentration of minority toponyms). On top of the interpolation result, circular dots are placed using the same color scheme; majority toponyms are represented in blue and minority toponyms in red.

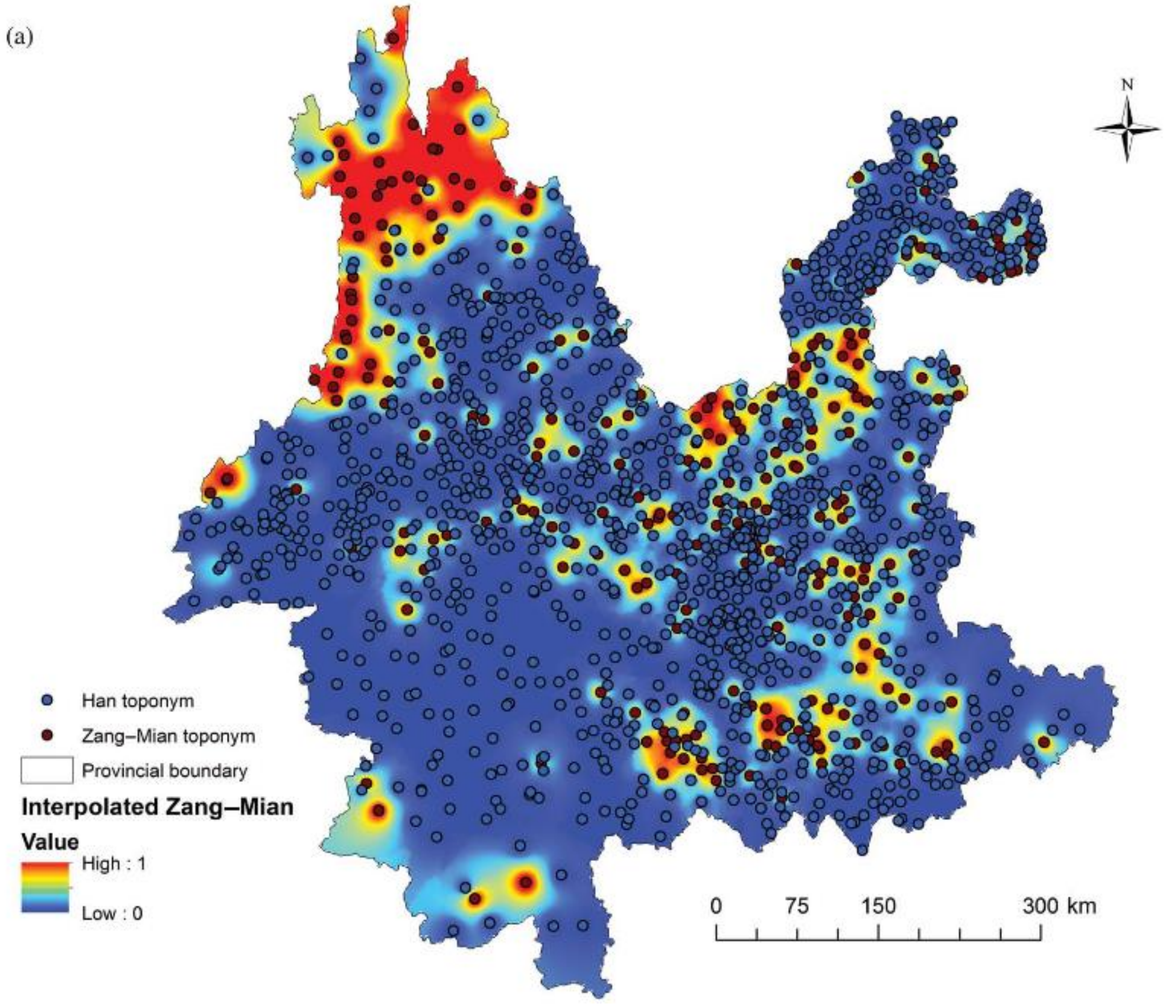

Figure 11 IDW result visualises relative concentration of Zang-Mian minority toponym group compared to Han majority toponym group (F. Wang et al., 2014, p.92 Figure 5.a)

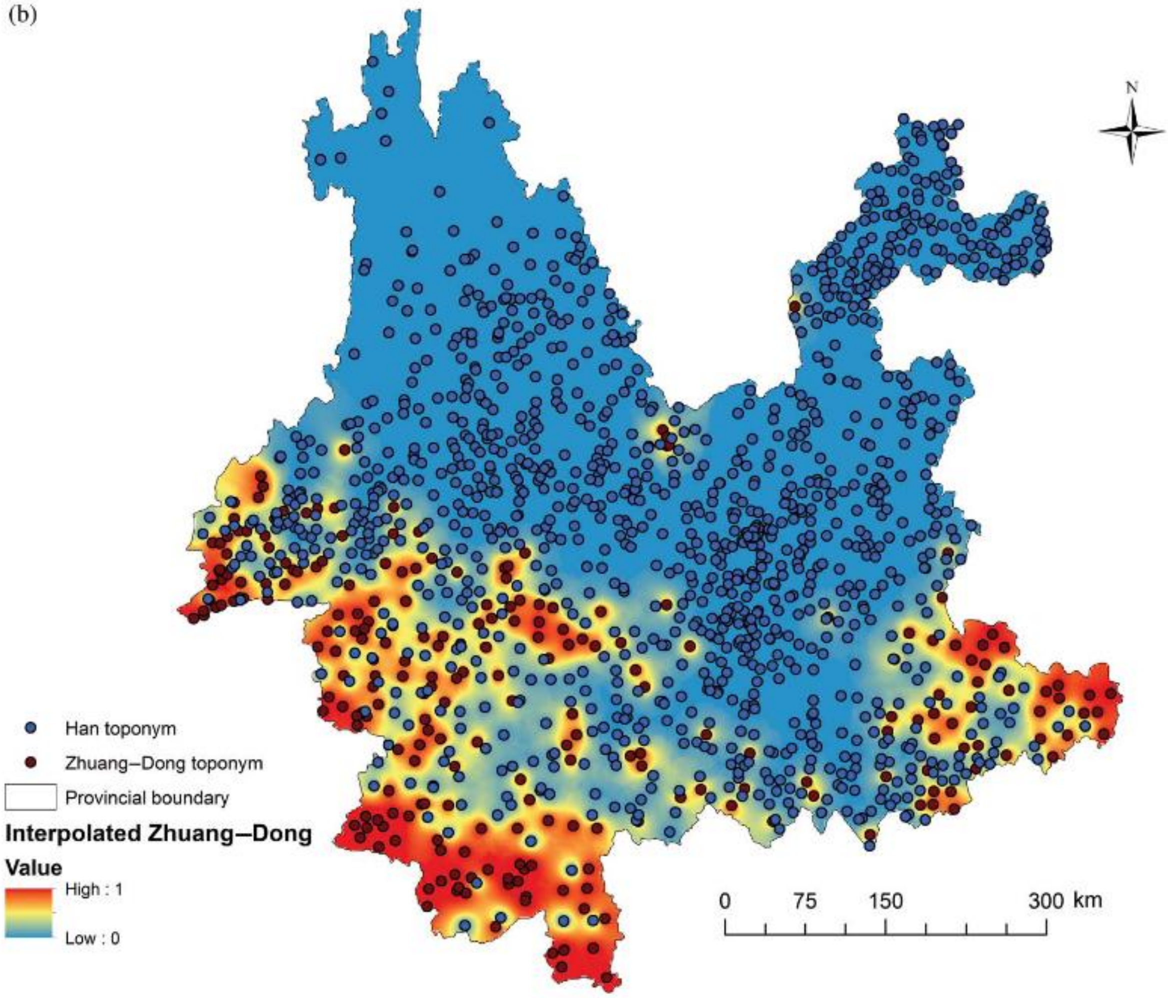

Figure 12 IDW result visualises relative concentration of Zhuang-Dong minority toponym group compared to Han majority toponym group (F. Wang et al., 2014, p.92 Figure 5.b)

Furthermore, Wang et al. (2014) used spatial cluster analysis to detect clusters where the ratio of minority toponyms is significantly higher within a circular window than outside. The different clusters are visualised by circles in different colors and sizes. Such as in Figure 11 and Figure 12, dot mapping method is used to visualise the different toponymic groups; majority toponyms are represented as circular light-gray dots, and minority toponyms are emphasized by different color and shape (Figure 13 and Figure 14).

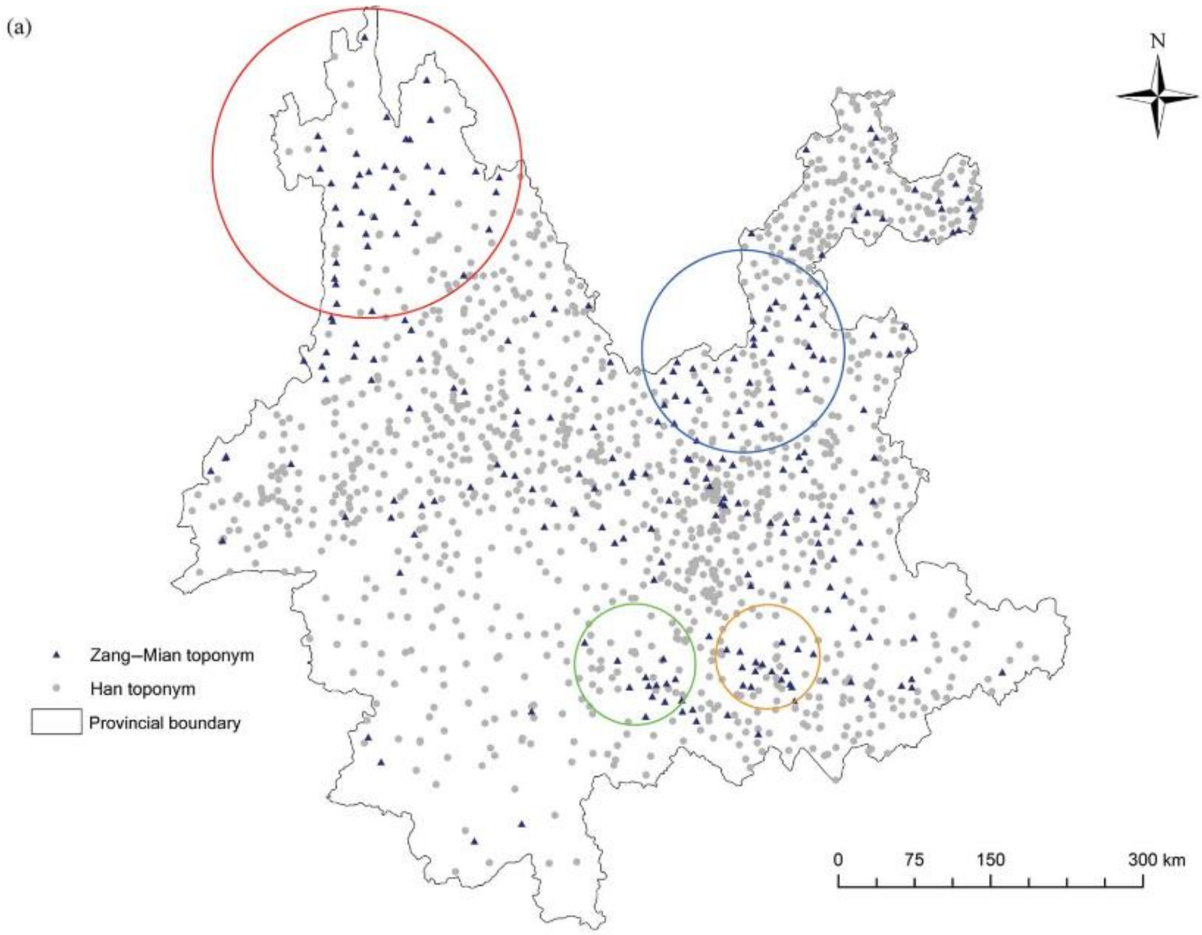

Figure 13 Spatial clusters of Zang-Mian minority toponyms (F. Wang et al., 2014, p.94 Figure 6.a)

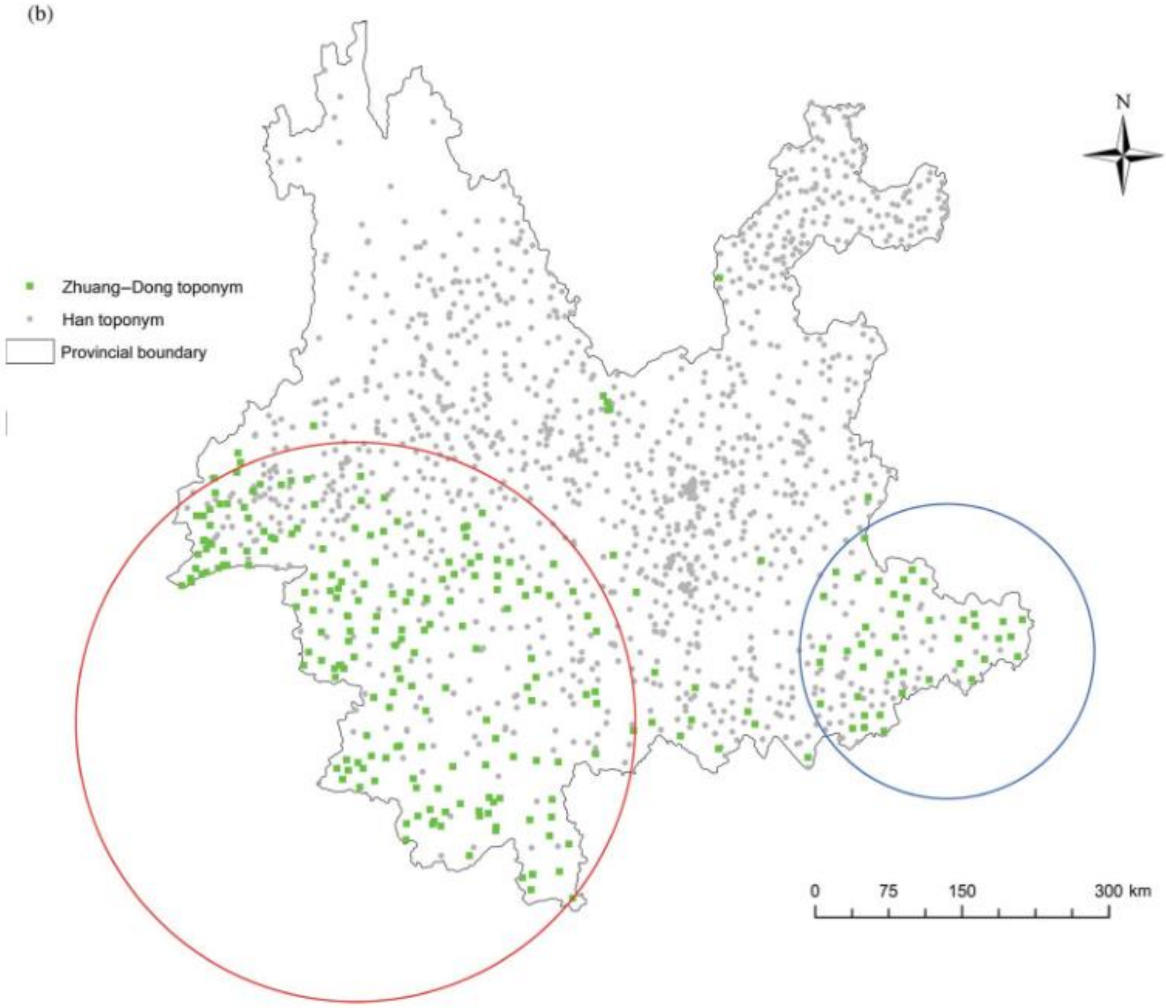

Figure 14 Spatial clusters of Zhuang-Dong minority toponyms (F. Wang et al., 2014, p.94 Figure 6.b)

For the comparison of toponym distributions and physical environment, Wang et al. (2012; 2014) used additional data. A digital elevation model (DEM) was used to compare toponyms to topography by extracting topographic variables, such as elevation, slope, and aspect (Figure 9) (F. Wang et al., 2012). Also, the researchers included the distance of toponyms from water streams, and railroads, and major roads (F. Wang et al., 2012, 2014). They calculated the average of all these variables for minority language toponyms and Han toponyms and calculated whether the difference between them was significant using a pooled t-test (F. Wang et al., 2012) and analysis of variance (ANOVA) (F. Wang et al., 2014). Furthermore, F. Wang et al. (2012) used land type data as Zhuang toponyms include morphemes with a meaning referring to land use or land cover and assigned the value 1 to toponyms on paddy or irrigated lands and value 0 to all other land types.

Fuchs (2015) applied KDE to toponym groups of different number and geographic extent. The researcher visualised the results as overlaid isopleths (Figure 15). The visualisation technique accounts for the overlapping of layers. The result shows the toponym density estimates among the different groups.

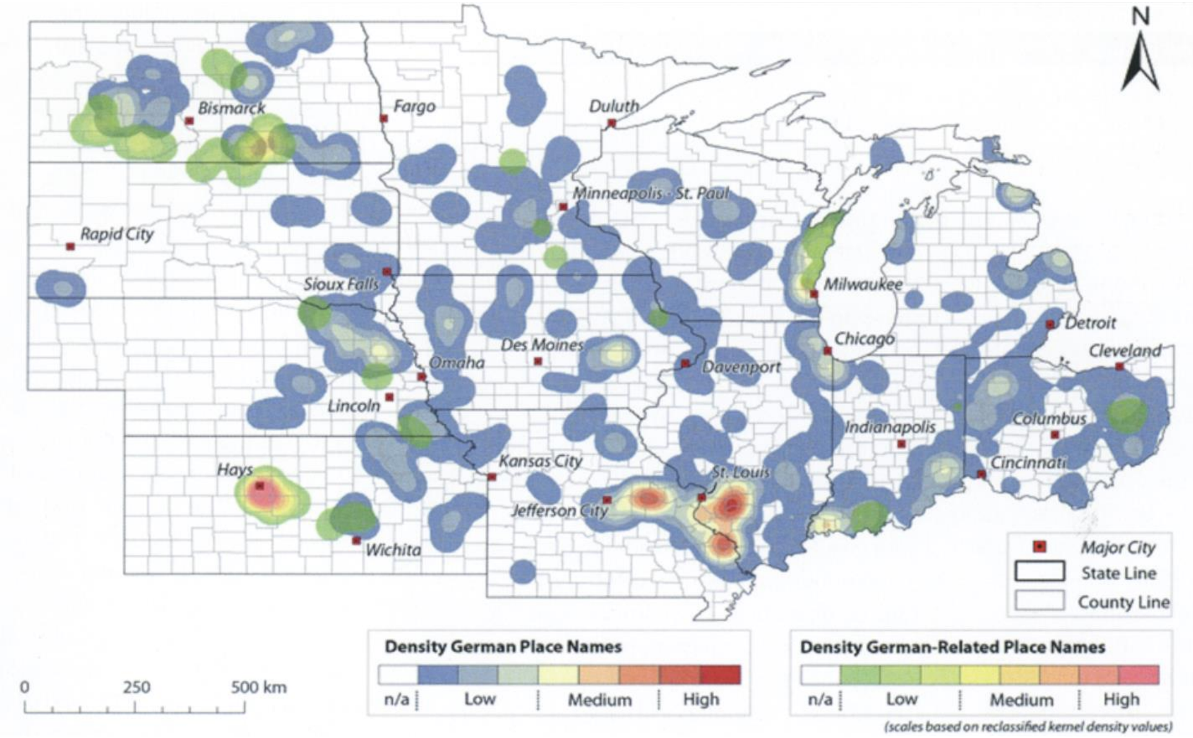

Figure 15 KDE result shows the density of German and German-related toponyms as overlapping isopleths (Fuchs, 2015, p.333 Fig.3)

Other research investigated the relationships of toponyms with ethnic groups, and of toponyms with the geographic environment in Guangdong province, China (Qian, Kang, & Wang, 2016; Qian, Kang, & Weng, 2016). The researchers classified the toponymic landscape by creating subsets of toponyms with different linguistic origin. Within each linguistic layer, toponyms were sub-grouped into their meaning with regards to the geographic environment (Figure 16). Qian, Kang, & Wang (2016) estimated the kernel density to analyse spatial concentrations of the toponymic layers. The results are visualised as continuous toponym density layers using the isopleth mapping technique (Figure 17). They reveal spatial patterns of different ethnic groups. Furthermore, Qian, Kang, & Weng (2016) created a series of thematic maps that show migration trends of different ethnic groups based on toponym analysis. An example is given in

Figure 18 showing the change of Zhuang population using isopleth mapping technique to visualise the continuous distribution of Zhuang people and flow mapping technique to indicate the direction of Zhuang migration. The so-called *toponym mapping method* applied by Qian, Kang, & Weng (2016) considers toponyms as original thematic data and uses additional data, such as census data.

General name	Meaning	Total counts
Zhuang toponymic layer		
na	Rice field	565
tong	Plains among mountains	353
Cantonese toponymic layer		
li	River	106
yong	Water channel	103
ji	Dike	113
sha	Coastal plains	281
chong	Cave	1608
Hoklo toponymic layer		
cuo	Settlements	170
yang	Waters	306
qian	Neighboring	28
Hakka toponymic layer		
zhang	Peak	156
she	Cultivated land	14
duan	Flat ground	183
bei	Mountain	1213
lang	Hillside	384

Figure 16 Counts of toponyms with a generic meaning ordered by linguistic toponymic layer (Qian, Kang, & Wang, 2016, p.166 Table 1)

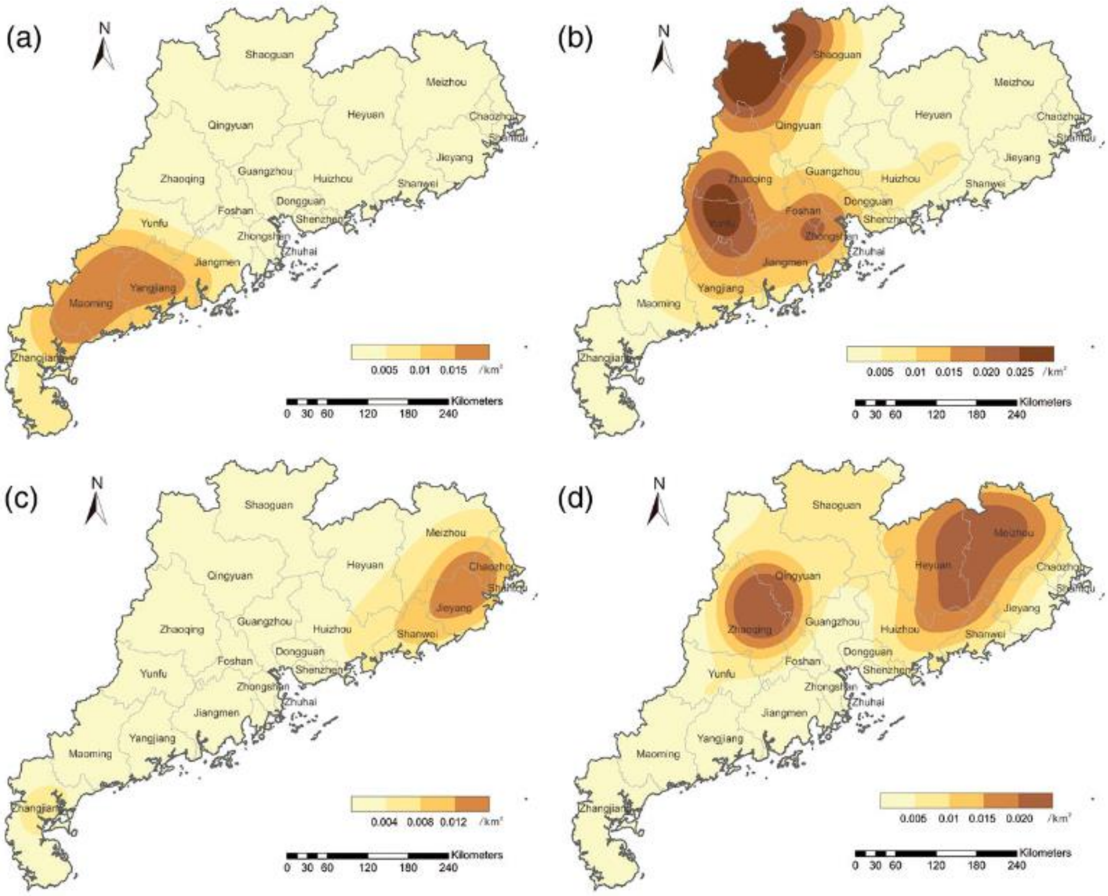

Figure 17 KDE result shows ccontinuous surfaces of different toponymic layers visualised as isopleths: (a) Zhuang toponymic layer; (b) Cantonese toponymic layer; (c) Hoklo toponymic layer; (d) Hakka toponymic layer (Qian, Kang, & Wang, 2016, p.167 Figure 2)

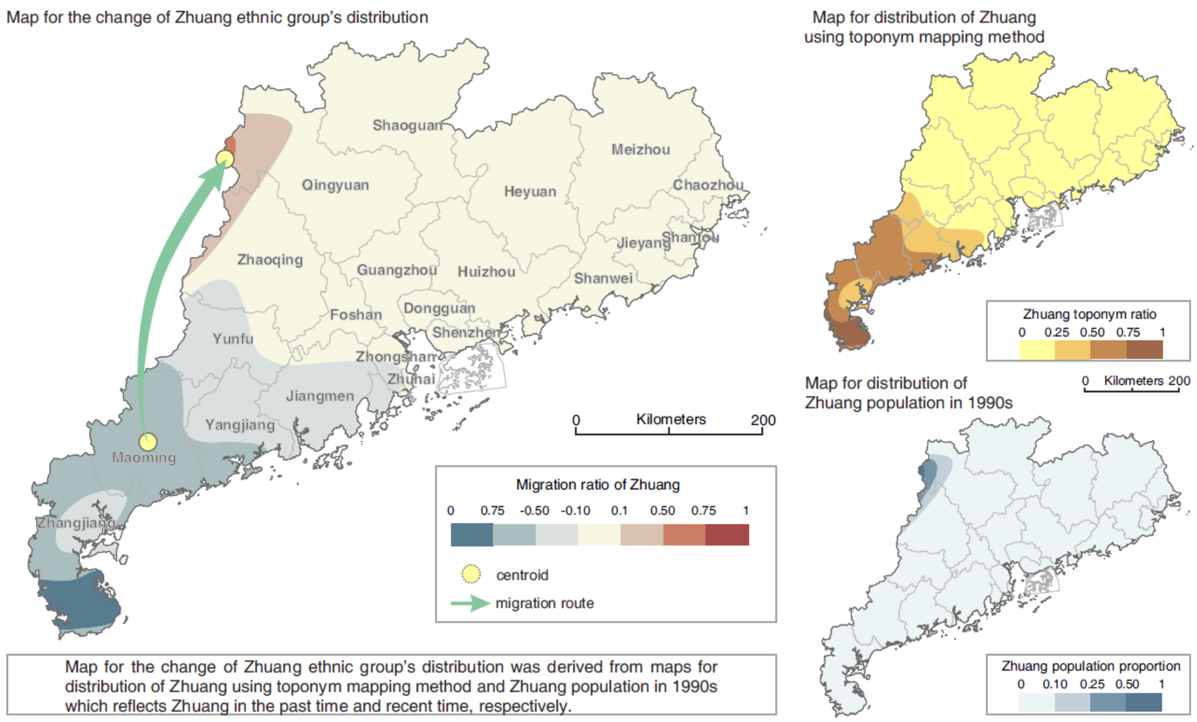

Figure 18 Toponym mapping method result visualises changes of Zhuang distribution (Qian, Kang, & Weng, 2016 Supplemental material)

The research by Zeini et al. (2018) used spatial cluster analysis and KDE to identify and visualise distribution patterns of specific toponym types, e.g., settlement-based toponyms (Figure 19), using their so-called *geographical-based approach* to classify toponyms (as described in 2.2.1.3). The KDE result is visualised using the isopleth mapping method; circles represent toponym clusters and individual toponyms are visualised as dots.

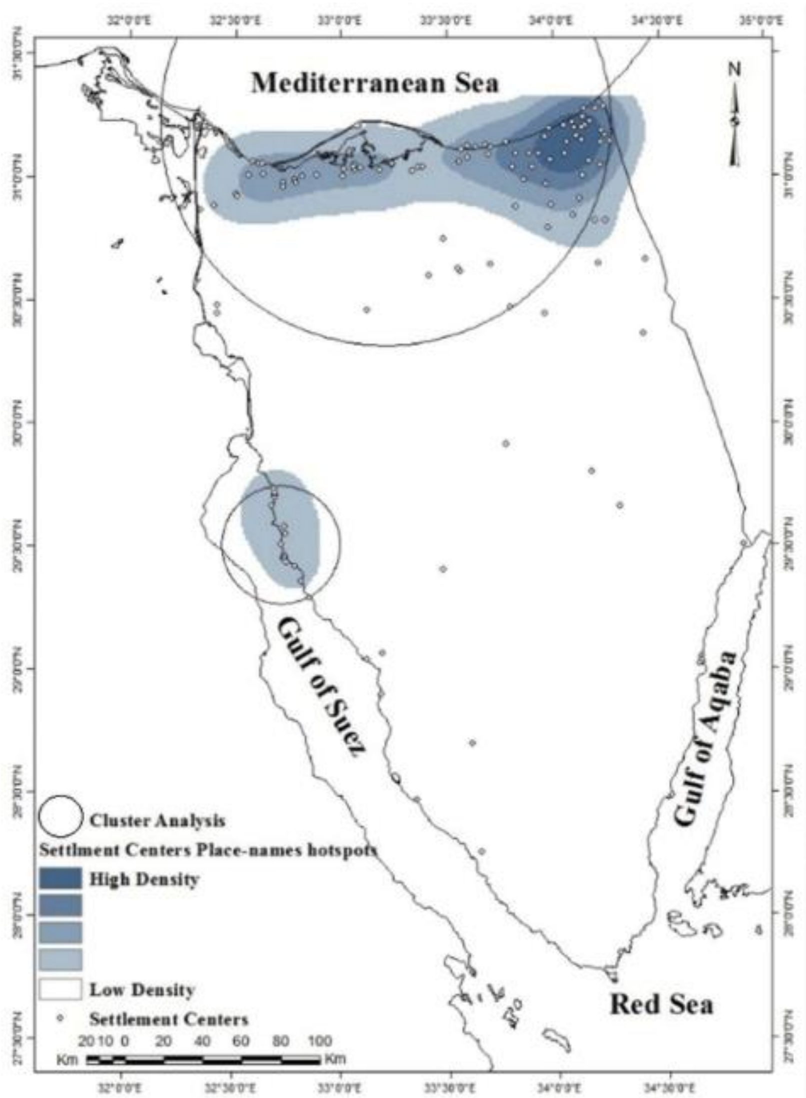

Figure 19 Distribution patterns of settlement-based toponyms (Zeini et al., 2018, p.188 Figure 7)

The research by Chloupek (2018) groups toponyms in Nebraska into nine different thematic categories, including the category of Indigenous toponyms. Nearest neighbour tests identified significantly clustered and dispersed distributions for toponyms created in a defined time range for three categories. These distribution patterns are presented visualising the KDE result as isopleths; dots represent the toponyms of the analysed group (Figure 20). Furthermore, other spatial patterns were detected, namely that Indigenous toponyms appear closely related to major rivers (Chloupek, 2018). However, toponyms from other categories, such as place names borrowed from the eastern United States, the classical world, or directly from Europe (with exception of toponyms of the natural environment), yield not to be related to their geographic environment or to the geographical feature they represent (Chloupek, 2018).

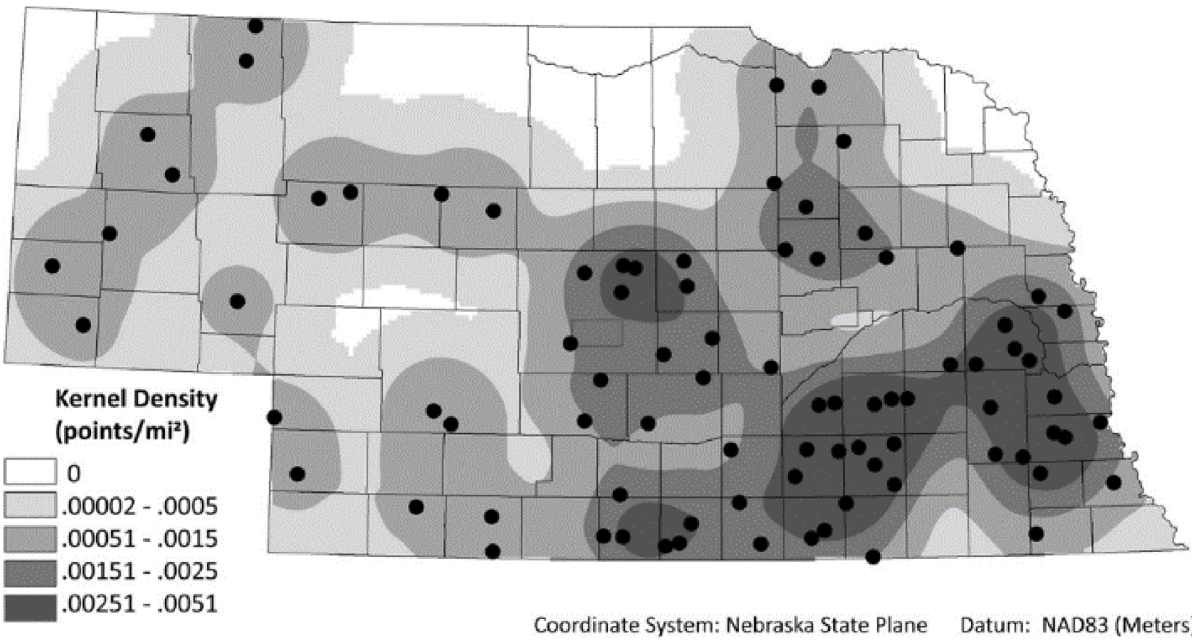

Figure 20 KDE result of toponyms of the category "place names borrowed from eastern United States" visualised using the isopleth mapping method (Chloupek, 2018, p.36 Figure 3)

The three standard cartographic symbol types in thematic mapping – also termed as geometric primitives – point, line, and area (polygon) symbols, are used to represent spatial phenomena on a map (Dent et al., 2009). The geometric primitives used in the selected toponymy research are mainly point and area symbols. In the outlined work, points vary particularly in shape, color, and size, and represent individual toponym locations. Area symbols are mostly used in mapping the KDE results by isopleths and vary in color. Used to less extent, line symbols serve as spatial reference in the outlined work; they represent the delimitations of the study areas.

To conclude, the GIS-based methods commonly used in the toponymic work outlined in this section are summarised. Spatial cluster analysis (e.g., Chloupek, 2018; F. Wang et al., 2014; Zeini et al., 2018), KDE (e.g., Chloupek, 2018; Fuchs, 2015; Luo et al., 2010; Qian, Kang, & Wang, 2016; Zeini et al., 2018), spatial interpolation methods, such as trend surface analysis or IDW (e.g., F. Wang et al., 2012, 2014), and statistical methods to test or calculate e.g., the distance of toponyms to rivers (e.g., Luo et al., 2010; F.

Wang et al., 2012, 2014) are commonly used methods to analyse toponymic distributions. To visualise toponyms individually, the dot mapping method is commonly used (e.g., Chloupek, 2018; F. Wang et al., 2012, 2014; Zeini et al., 2018). Isopleth mapping often visualises the results obtained from KDE (e.g., Chloupek, 2018; Fuchs, 2015; Qian, Kang, & Wang, 2016; F. Wang et al., 2012; Zeini et al., 2018). Other thematic mapping techniques, such as proportional symbol mapping (e.g., F. Wang et al., 2012) or flow mapping (e.g., Qian, Kang, & Weng, 2016) are used to less extent in the selected research.

2.2.4. GIS-based Methods in Toponymy Science

This section presents spatial analysis and cartographic visualisation methods that are commonly applied to toponym distributions in GIS-based research. Some of the methods used in the toponymy research outlined in the previous sections are discussed in more detail.

2.2.4.1. Dot Maps and Dot Density Maps

Toponyms in this research are treated as points in space. Consequently, to analyse toponym distributions means to find point patterns that can be visualised on a map. In terms of mapping methods, there are a variety of methods used to map points or their densities; dot mapping represents one of them.

Dot maps are characterized by dots that are placed at each data location with the same point symbol; each data point is represented by exactly one dot (O'Sullivan & Unwin, 2010). Design consideration for this *one-to-one mapping* approach is therefore limited to size, shape, and hue of the point symbol (O'Sullivan & Unwin, 2010). The dot size should be selected so that the symbols are large enough to be visible on the map and small enough to be distinguished from other dots (O'Sullivan & Unwin, 2010). Examples of dot maps using *one-to-one mapping* approach are shown in Figure 6 & Figure 7. In a *many-to-one mapping* approach, data points are aggregated and presented as one dot. This approach is useful to improve the overall visual impression which can be deteriorated in *one-to-one mapping* due to different perceptions of the density of dots caused by different spaces between dots and dot coalescence (O'Sullivan & Unwin, 2010). The result of a *many-to-one mapping* approach is therefore called *dot density map* whose design constraints include the selection of the per-dot value. This value can either be chosen in a way that dot coalescence only occurs in densest areas or that the original count can still be retrieved from the map by making each dot countable and distinct (O'Sullivan & Unwin, 2010). Dot density maps are not to be confused with proportional symbol maps. Dot density maps use one dot to represent the count of several data points, while proportional symbol maps scale the dot in proportion to the magnitude of the data occurring at a single point location (Slocum et al., 2013).

In conclusion, there is two approaches of mapping dots: the *one-to-one mapping* approach creates one dot for each data point resulting in *a dot map*, whereas the *many-to-one mapping* approach resulting in a *dot density map* aggregates data points and represents them as one dot.

2.2.4.2. Georeferencing Historical Maps

The georeferencing of maps is not covered in section 2.2.3. However, this research uses georeferencing to analyse the relationship of toponymic distributions with language areas extracted from a historical map. The method is outlined in this section.

Georeferencing refers to the alignment of a digital image of a map with a digital reference map containing accurate coordinates and serves the intention to transfer map features to a GIS. (Clark & MacFadyen, 2020). The reference map can be a base map which is usually provided by GIS software. The digital image of the map might be provided in raster format. According to Király et al. (2008), there is two methods to transform the raster image to its real-world coordinates: (1) based on parameters, and (2) based on control points. This research uses the second transformation method. By adding two control points, the map is scaled and rotated; however, more control points are necessary to improve accurate georeferencing, for example at the four corners of the map (Geospatial Historian, 2013). Adding control points in areas where accuracy on the map is especially concerned is good practice (Geospatial Historian, 2013). Geographical features that are represented on both maps can help to locate suitable control points (Clark & MacFadyen, 2020). Therefore, the base map should be selected according to the geographical features represented on the map to be georeferenced. Having selected control points, their accuracy should be assessed. According to Clark & MacFadyen (2020), there is two main assessment methods: (1) the residual error assesses the accuracy of individual control points – the higher the value of the residual error, the larger the distance between the control point location in the georeferenced raster and the reference map – and (2) the root mean square error (RMSE) calculates the overall average discrepancy of the entire set of control points between raster and reference map. For both measures, it is important to consider scale. While a RMSE of ~ 1,000 meters is acceptable for a small-scale country map, it might not be accurate enough for georeferencing a large-scale city map (Clark & MacFadyen, 2020). Having assessed the accuracy, the output can be exported as a georeferenced raster containing the coordinates from the georeferencing process.

Having georeferenced a historical map it can be used as an additional information source included in toponym analysis. In a further step, geometric primitives can be digitized in order to apply geoprocessing tools and analyse spatial relationships between the toponymic data and the digitized features.

2.2.4.3. Kernel Density Estimation

This section outlines the theory of kernel density estimation (KDE) applied to point distributions and shows its application in toponym science.

KDE is an important method to analyse spatial patterns of point distributions (Zahtila & Knura, 2022). It transforms point patterns into estimated spatial densities (O'Sullivan & Unwin, 2010). The idea behind this method is that there is a density of points at any location within the study area – not only at point locations – that can be estimated (O'Sullivan & Unwin, 2010). Kernel density applied to point features in space uses a moving window, known as *kernel*, with a specified radius, known as *bandwidth*, to estimate the local density of points (O'Sullivan & Unwin, 2010; Y. Wang et al., 2019; Zahtila & Knura, 2022). The shape of the kernel is determined by the kernel (shape) function, which measures the distance decay effect of the points (Zahtila & Knura, 2022). The simplest approach uses a circular kernel (Figure 21) whereas more complex approaches use distance-weighted kernel functions, such as quartic kernel function (Figure 22), resulting in undulated kernel shapes (O'Sullivan & Unwin, 2010; Silverman, 1986; Zahtila & Knura, 2022).

The kernel density f at a location with spatial coordinates (x, y) is estimated as (adapted from Silverman, 1986; as cited in Zahtila & Kura, 2022, p.4):

$$\hat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{d_{i,(x,y)}}{h}\right),$$

where n is the total number of points, h is the bandwidth, $d_{i,(x, y)}$ is the distance between location i (where i are the number of points under concern) and location (x, y), and K is the kernel function (Zahtila & Knura, 2022, p.4). The number of points is counted within each kernel which is centred at the location where the estimate is to be made (O'Sullivan & Unwin, 2010). The bandwidth controls the amount of smoothing and is therefore also known as the smoothing parameter (Chen, 2017; Silverman, 1986). This means the higher the bandwidth, the smoother and more generalized the density. The smaller the bandwidth, the better local patterns are maintained (Zahtila & Knura, 2022). According to Zahtila & Knura (2022), it is generally accepted that the kernel function is not critical to the result, whereas the bandwidth is.

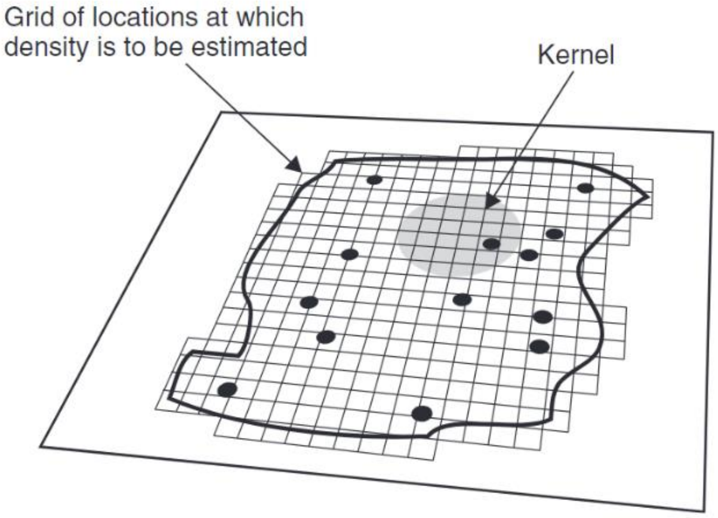

Figure 21 KDE using simple, or naïve, kernel function (O'Sullivan & Unwin, 2010, p.69 Figure 3.2)

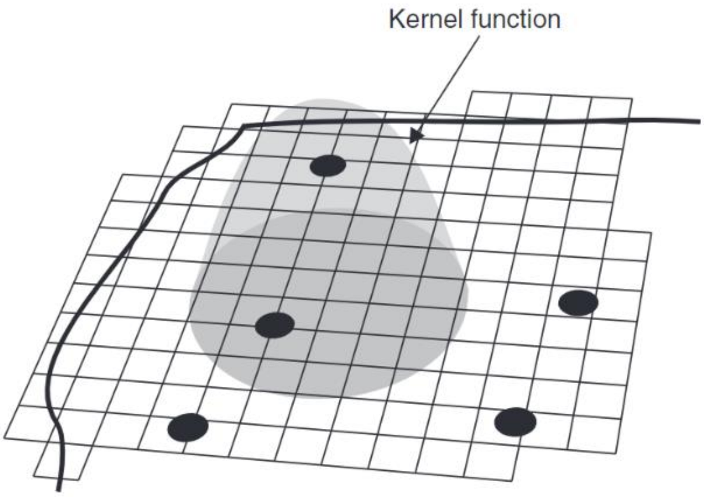

Figure 22 KDE using quartic distance-weighted kernel function (O'Sullivan & Unwin, 2010, p.69 Figure 3.3)

The KDE tool in a GIS software generates a raster output for which each cell contains a density value calculated by adding the values of all kernel surfaces that overlay the centre of the raster cell (Figure 23) (O'Sullivan & Unwin, 2010). Consequently, every cell where the kernel covers at least one point contains a density value. The kernel density value is 0 for cells where no point is covered by any kernel. To visualise the results of the KDE, the cell density values can be assigned to a defined class. The class breaks are defined by choosing a suitable classification method, such as Jenks natural breaks for non-uniform distributions creating classes of irregular class widths including varying frequency of observations per class (Chloupek, 2018). The importance of data classification methods and the number of classes is discussed in 2.2.4.6.

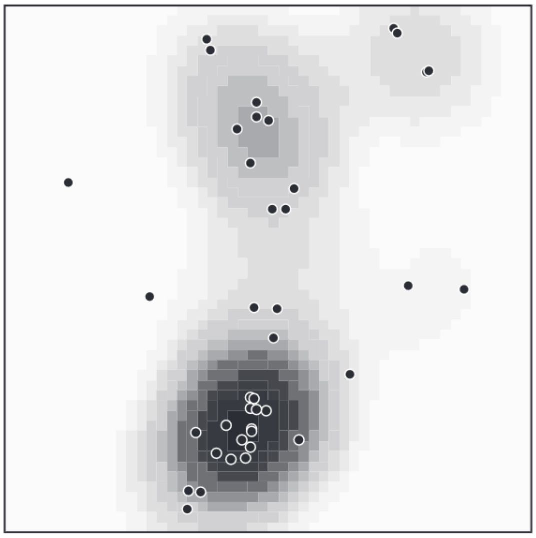

Figure 23 Typical output surface from KDE and its original point pattern (O'Sullivan & Unwin, 2010, p.70 Figure 3.4)

KDE applied to toponymic data allows to detect high and low concentrations of toponyms in space (Fuchs, 2015). There are many examples of extensive toponymy research using KDE: Chloupek (2018) see Figure 20; Fuchs (2015) see Figure 15; Luo et al. (2010) see Figure 8; Qian, Kang, & Wang (2016) see Figure 17; Zeini et al. (2018) see Figure 19. Furthermore, KDE results are commonly visualised using the isopleth mapping method covered in the next section.

2.2.4.4. Isopleth Mapping

As identified in section 2.2.3, isopleth mapping is a common method in toponymy research to visualise KDE results. This section presents the mapping method more in detail.

In general, isopleth maps aim to simplify information about a region by showing areas with continuous normalized distribution. Isopleth mapping is – together with isometric mapping – a category of isarithmic mapping that is defined as a planimetric graphic representation of the surface (Dent et al., 2009). An isarithmic map uses isolines to symbolise quantitative values as undulating continuous surfaces. To map data isarithmically, the mapped geographic phenomenon is assumed to occur everywhere in space. Therefore, isarithmic mapping method is usually used to map variables, such as temperature, elevation, or precipitation. Point data used for isarithmic maps are either of true or conceptual nature: true point data are measured at a location, while conceptual point data are considered to be located at points; however, they are collected over areas (Slocum et al., 2013). Isopleth mapping uses conceptual point data occurring over geographic areas (Dent et al., 2009). Isopleth mapping requires to normalise the data considering the area over which the conceptual point data is collected (Dent et al., 2009). Therefore, the values inside the delimited areas with similar regional aspects – also known as isopleths – are normalized using the areal magnitude of the isopleth to calculate the ratio. According to Dent et al. (2009) In isopleth mapping, lines or colors are used to define isopleths. The shape of the isopleths is driven by the underlying data and not by pre-defined areas, such as census units in choropleth maps (Dent et al., 2009).

Applied to toponyms, isopleth mapping can be used to show toponym densities as a continuous surface. Toponymic studies have used this mapping method to visualise results obtained from kernel density estimation (see example research given at the end of section 2.2.4.3 along with references leading to figures in section 2.2.3). For this approach, the visualisation often uses hypsometric tints (shaded areas) for which light and dark tints are associated with low and high concentrations (Slocum et al., 2013).

2.2.4.5. Hexagon Mapping

This section outlines the hexagon mapping method as an alternative to the visualisation of toponym density estimates using KDE and isopleth mapping. Whereas KDE estimates toponym densities, the hexagon mapping method aggregates toponyms into hexagonal grid cells. The toponym densities are visualised within the hexagons as classified toponym counts.

Hexagon maps use regular shaped hexagons to aggregate data. This thematic mapping method is a good alternative to using polygons with artificial boundaries, such as administrative borders. One constraint that comes along with using irregular shaped areas for thematic mapping is what Openshaw (1983) defined as the modifiable areal unit problem (MAUP). This constraint refers to the aggregation of geographic data from a detailed level within areas of varying sizes and shapes. Using irregular areas to map thematic data produces wrong perceptions on distribution patterns (see Figure 24). Aggregating data into regular shaped polygons, such as hexagons, allows to normalize data and therefore to avoid these misperceptions (Esri, 2022; McKenzie, 2022).

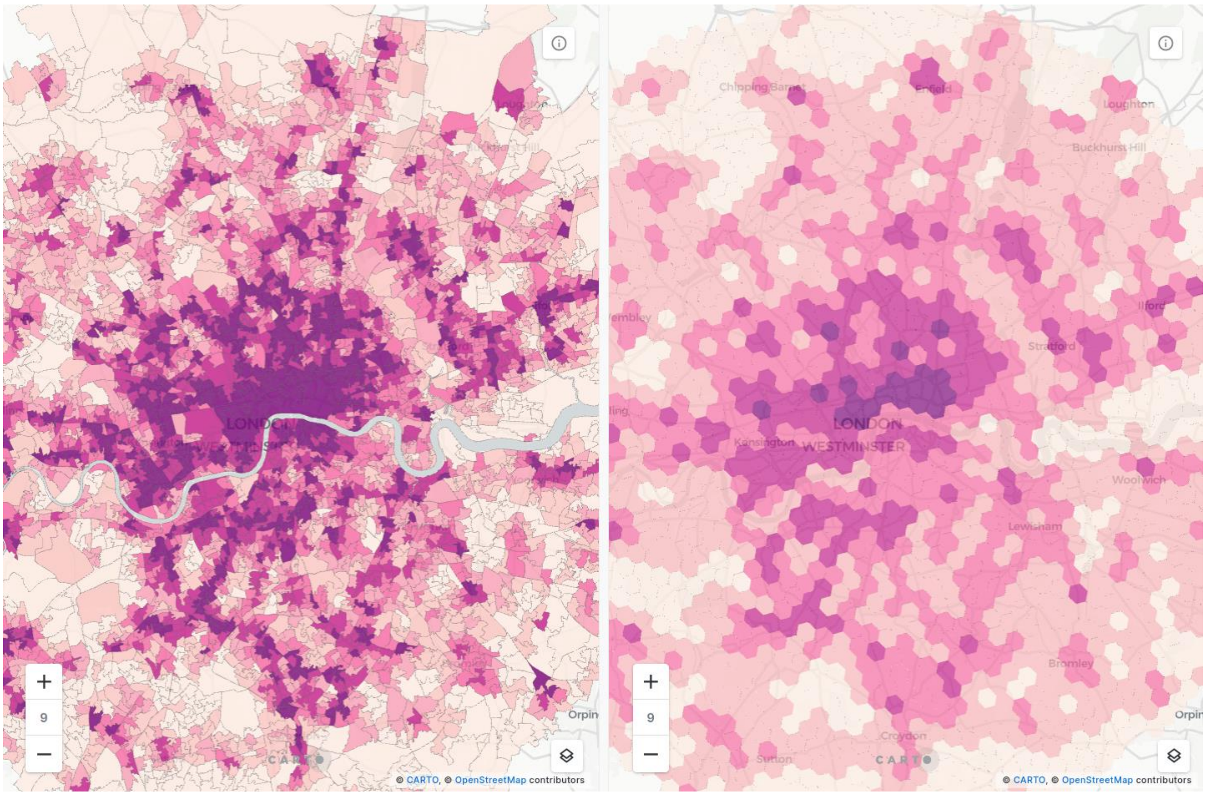

Figure 24 Comparison of thematic map using administrative borders (left) and hexagons (right) (McKenzie, 2022)

Before the times of GIS-based cartography, hexagon maps had been very labour intensive to produce and therefore rather hard to find. However, the historical map by John Leighton (Figure 25) uses hexagons color-coded according to their proximity to the centre of London (Leighton, 1895). For comparison, Leighton (1895) created a map of London showing London and its boroughs as irregular shapes (Figure 25). Leighton (1895) suggests this mapping method including an underlying systematic approach with the aim to structure the city and improve navigation.

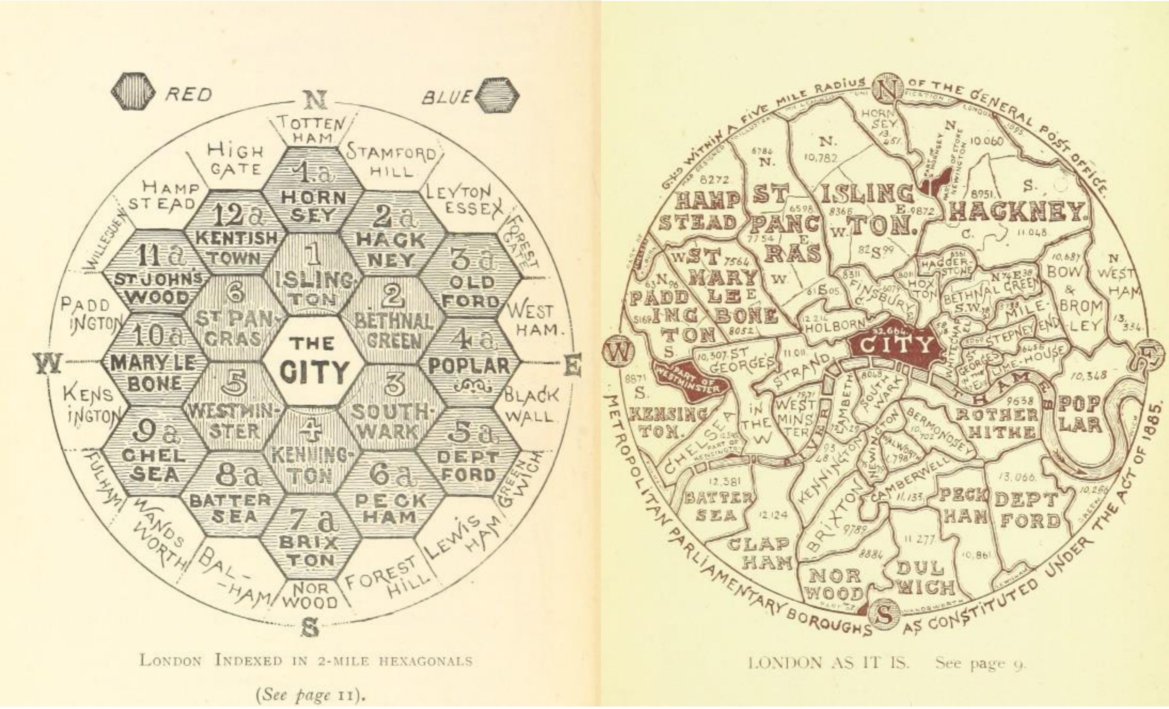

Figure 25 Historical hexagonal map of London (left) and historical map of London "as it is" (Leighton, 1895, p.9 & 11)

Such as hexagons, squares and triangles also generate a tessellated regular grid. However, according to McKenzie (2022) , some properties make the use of hexagons superior to the use of other regular grid shapes. For example, the distance of a hexagon's centroid is the same to all neighbouring centroids. This concept of central place theory is also used by Leighton (1895) (Figure 25) by representing and surrounding London's city centre by hexagons. Also, other shapes with sharp angles are more prone to outliers located at the shape's angles. Due to their shape that is closer to a circle than squares or triangles, they possess the ability to smooth distributions. Therefore, hexagons outperform other shapes at representing gradual changes in space and linear distributions.

Hexagon mapping can also help to tackle the issue of overlapping data points. By aggregating points into hexagons and symbolizing the hexagons with graduated colours, each hexagon will represent the total count of points within it. This way, the number of points at a location is not underestimated due to overlapping.

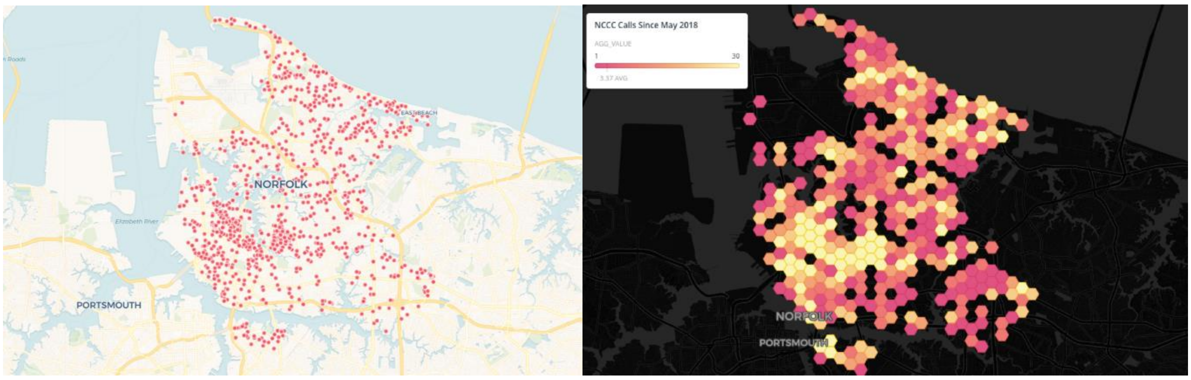

Figure 26 Avoiding overlapping data points by aggregating points into hexagons (Mongon, 2018)

2.2.4.6. Data Classification

This section deals with finding suitable methods to classify quantitative data.

The classification method has a large impact on the map's appearance as shown in Figure 27 where the choropleth map on the left uses equal intervals and the map on the right uses quantiles to classify the percentage of population of Asian ethnicity in several census area units in North Shore City, Auckland, New Zealand in 2006 (O'Sullivan & Unwin, 2010). This research suggests using the natural breaks method to classify data. This method uses logical and natural grouping of the data with the aim *"to minimize differences between data values in the same class and to maximize differences between classes"* (Slocum et al., 2013, p.72).

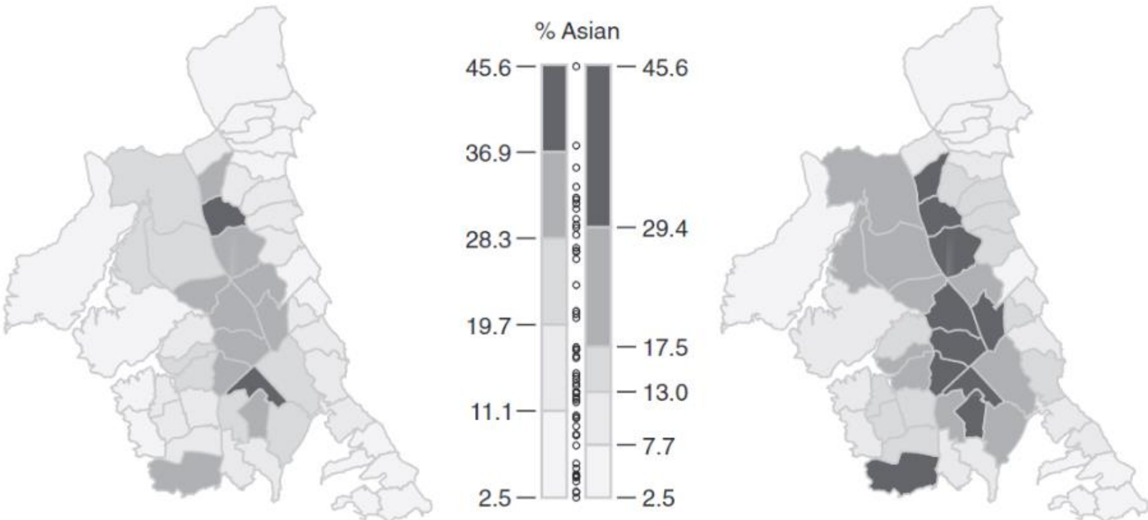

Figure 27 The effects of classification method on the appearance of a choropleth map (O'Sullivan & Unwin, 2010, p.75)

Not only how the class ranges are set, but also the number of classes has to be considered. How many classes should be used for quantitative numerical data, is outlined by Dent et al. (2009) who suggest selecting the number of classes based on the formula by Sturges (1926; as cited in Dent et al., 2009):

$$C = 1 + 3.3 * \log(n)$$

where *C* is the number of classes and *n* the number of data points. However, Dent et al. (2009) state that the result should only be used as a rule of thumb. According to the formula above, with a number of 200 data points (*n*) we would already obtain 9 classes which might cause the map's content to be unclear and less distinctive. Also, considering large data volumes to be investigated in particularly quantitative and extensive research this formula seems to be outdated. According to Dent et al. (2009), it is therefore the art of cartography to choose a suitable class number which involves the cartographer to have an understanding of the topic and the nature of the data; most cartographers stick O'Sullivan & Unwin (2010)sses. O'Sullivan & Unwin (2010) state that five to seven classes is appropriate.

In conclusion, how we classify our data has a large impact on the appearance of a map. Choosing a suitable classification method and number of classes is largely dependent on the topic and the data to be classified. However, choosing between four to seven classes is a common practice in map-making (Dent et al., 2009; O'Sullivan & Unwin, 2010).

2.2.5. Semi-Structured Interviews

As toponymy research is interdisciplinary, this research used semi-structured interviews to assess the results of this research with experts from scientific fields other than cartography or GIS-based research.

A semi-structured interview is a qualitative data collection method that is characterized by its flexibility (Edwards & Holland, 2013). As preparation for the interview, the researcher prepares a list of questions or topics to discuss (Edwards & Holland, 2013). However, it is not predetermined how the questions are exactly formulated, at what stage of the interview the questions are asked, or how the interviewee can respond to them (Edwards & Holland, 2013). This flexibility opens the possibility for a dialogue or discussion as the interviewee can also ask questions to the interviewer (Edwards & Holland, 2013). Nevertheless, by giving every interviewee the same theoretical framework, it makes the interviews comparable (George, 2022). However, the flexibility of a semi-structured interview allows the interviewee to shape the development of the interview. This makes this interview type useful to focus on different aspects of a thematic (George, 2022).

3. Methodology

This chapter outlines the research design in detail with the aim to make it replicable for similar purposes using a different study area and different datasets. It is divided into three sections. The first section describes the preparation necessary to carry out the research. The second part addresses the methodology used to develop a toponym classification system. The last section outlines the GIS-based analysis and visualisation methods applied to the classified toponymic data and other datasets.

3.1. Research Preparation

This section deals with the essential steps to prepare the research.

The research preparation is divided into five parts: (3.1.1) defining the study area and finding a toponymic dataset that covers this area; (3.1.2) conducting study-area-specific literature analysis with regards to toponymy and morphemes influential in Indigenous-derived toponymy; (3.1.3) selecting additional datasets of environmental variables, historical and current language areas, and political boundaries; (3.1.4) considering technical requirements; and (3.1.5) preparing the toponymic data for analysis.

3.1.1. Study Area and Toponymic Data

The research preparation starts with defining the study area and finding a toponymic dataset for this area that includes the following information[1]:

1. Spatial information, e.g., longitude (*long*) and latitude (*lat*)
2. Toponym, preferably stored in different manners as suggested in the following:
    - Reading order; upper-cased; without spaces, hyphens, or diacritics (*versionI*)
    - Reading order, without diacritics (*versionII*)
    - Reversed reading order; upper-cased; without spaces, hyphens, or diacritics (*versionIII*)

[1] The exemplary column name in parenthesis is used again in Table 5 in section 3.2.4.

3. Information on geographical *feature type* or *feature class*, e.g., populated place, hydrographic feature, etc. (*fc*)

Datasets that store toponyms in different manners facilitate querying. Storing toponyms in reading and reversed order is useful for toponyms that consist of two or more words. For example, toponyms, comprised of two words, are queried for a particular ending (suffix) in either the first or the second word. Furthermore, spaces, hyphens, or diacritics might impede accurate toponym queries. Therefore, this research suggests using a dataset that stores the toponyms without spaces, hyphens, or diacritics. However, datasets storing the data differently than suggested in this research can also be used. However, limitations in querying must be considered and queries must accordingly be adapted. Furthermore, information on geographical feature type or feature class is necessary to apply the suggested methodology. The feature type defines which geographical element is denominated by a toponym. For example, is the toponym given to a mountain, a river, or a city? The feature class would be an aggregation of different feature types into a broader class, e.g., a vegetation class that comprises features of the type forest, meadow, or grassland.

3.1.2. Literature Analysis

Further preparation regards literature analysis that focusses on answering the following questions:

Q1. What characterizes the toponymy of the study area, possibly linguistically and historically?
Q2. Which Indigenous languages are influential in the toponymy of the study area?
Q3. Which morphemes deriving from these Indigenous languages have a meaning referring to the geographic environment and occur frequently in the study area's toponymy?
Q4. Do these morphemes have variations?
Q5. Do these morphemes occur as prefixes, suffixes, or are their positions not indicated or irrelevant?
Q6. How *unique* are these morphemes?
Q7. Which generic meaning do these morphemes have?
Q8. To which feature type could the meaning of theses morphemes relate to?
Q9. Which environmental dataset could represent the meaning of these morphemes?

This research recommends answering these questions in the described order. However, the order can be adapted. For example, in the case that research is based upon pre-defined Indigenous languages or morphemes, Q2, Q3, and Q4 are already answered.

The steps necessary to answer these questions are described in detail in the following paragraph.

Analysing literature about the toponymy of the selected study area aims to collect information on (1) Indigenous languages with a high influence in the study area's toponymy and (2) morphemes that occur frequently in toponyms deriving from these languages. Furthermore, morpheme variations are collected; according to Lieber (2009, p.158), these are known as *allomorphs* and are defined as *"phonologically distinct variants of the same morpheme"*. For the data querying it is important to find out whether the morphemes occur as prefixes, suffixes, or whether their positions are not indicated or irrelevant (see section 2.2.1.2). The analysis also focusses on exploring the *uniqueness* of these morphemes; this means finding out whether morphemes might easily be confused with words or morphemes from other languages (see section 3.2.1). Also, the generic meanings of the selected morphemes are identified and associated with feature types. For example, toponyms carrying a morpheme with the generic meaning *mountain* are associated e.g., with the feature types *mountain, peak,* or *ridge*[2]. Furthermore, a dataset of an environmental variable is selected based on the generic meaning. For example, if the generic meaning of a morpheme is *river*, a dataset of the study area's hydrography is selected. In the end, the morphemes are compiled including their linguistic origins, allomorphs, generic meanings, literature references, associated feature types, and environmental variables represented by a dataset. The compilation does not aim for completeness but represents a selection of morphemes of interest which are included in the classification and GIS-based analysis and visualisation. The findings from literature analysis are structured and compiled in form of a table, for example. However, it is advisable to note down all kind of findings – along with their references.

3.1.3. Additional Data

In the third part of the research preparation, additional datasets are selected. The data should provide information on the following topics:

- Historical Indigenous language distributions
- Current Indigenous language distributions
- Different datasets on environmental variables, e.g., hydrography, elevation
- Boundary layers for different levels depending on scale and extent of the study area, e.g., country borders, state borders, or municipality borders
- Census data

[2] However, this does not imply that a toponym with a generic meaning is classified as the feature type it is associated with based on its generic meaning.

Data on historical and current Indigenous language distributions is necessary to analyse the spatial relationship of toponyms with language areas. For the historical distributions, the data source can be a digital historical map depicting language areas as polygons – as it is the case in this research. When using a historical map, the map must be available in good resolution permitting to distinguish between different language areas. Furthermore, the map must be georeferenced as outlined in section 3.3.4. Datasets on environmental variables are selected according to the findings from literature analysis outlined in section 3.1.2. They depend on the generic meanings of the selected morphemes and are used to analyse the spatial relationship of toponyms with their geographic environment. Furthermore, it is suggested to use boundary layers to provide spatial reference in the map visualizations. Considering that this research is extensive in nature and takes place on a small scale, it is advisable to use boundary layers. However, they are optional. Census data is used in the dataset exploration step in section 3.1.5.2 to calculate toponym densities in relation to the population within a census area.

3.1.4. Technical Requirements

This section addresses technical requirements that must be considered to carry out this research.

It is crucial to adequately choose software, tools, and programming language as it strongly determines the analysis possibilities and research outcome. There is no *correct* software, tools, or programming language for the different analysis steps, but the choice depends mainly on the researcher's knowledge and software availability. For data preparation, toponym classification, and GIS-based analysis and visualisation, different programs can be used. Free and open-source programming languages, e.g., R or Python, along with their open-source software, e.g., RStudio or Spyder, are useful for data preparation. For the toponym classification and querying, many open-source database management systems using different programming languages can be used. ArcGIS Pro and QGIS are examples of GIS software. While QGIS is open-source, ArcGIS Pro requires a software license.

3.1.5. Toponymic Data Preparation

This section and the last step of the research preparation covers the preparation of the toponymic data. It involves data modification, the exploration of the dataset, and the integration of the toponymic data into a database.

First, the attribute information stored in the toponymic dataset must be comprehended. A look into the metadata might help to understand the data structure, how the data is encrypted, and definitions. Loading the data in the form of a table into a software, such as RStudio or Spyder, might facilitate this step and the following ones. The exploration should answer the following questions (not necessarily in this order):

Q1. Which attribute information does the data store?
Q2. Which attribute information is essential for the research tasks?
Q3. In which format is the spatial information provided?
Q4. How is the toponymic information stored?
Q5. How are feature types defined and possibly grouped into feature type classes?[3]

3.1.5.1. Data Modification

After having understood the dataset, its structure, and its attribute information, the data is modified.

First, missing values (NA-values) might need to be handled. This step avoids complications during the data processing, which includes querying and updating the data. One way to handle NA-values is to replace them by the value 0. Also, other attribute information not necessary for the analysis is excluded. For this step, it serves to generate a copy of the original data and to modify the copy. The attributes that are essential for this research are mentioned in section 3.1.1.

3.1.5.2. Dataset Exploration

The following paragraph outlines the steps carried out to explore the dataset and to get an overview of toponym concentrations and densities (density estimates respectively) by creating first data visualisations.

The pre-processed toponymic dataset, census data of the study area, and optionally boundary layers are imported into GIS to explore the overall distribution of toponyms in the study area. There are two exploration approaches: (1) kernel density estimation is applied to the toponymic data and the result is visualised as isopleths showing high and low estimated toponym concentrations within the whole study area, (2) relative toponym densities on a municipality, state, or other level (depending on available data and scale of study area) are calculated dividing the toponym count by the population within an area. The second approach uses the population extracted from the census data and visualises the result as a choropleth map using graduated color values. Classifying the densities by using color value is suitable for quantitative data. For the visualisation in both approaches, a boundary layer is added to enhance orientation by providing additional spatial reference.

[3] This question is particularly relevant for the section 3.2.

This exploration step is useful for the discussion of the results and the derivation of hypotheses concerning toponym distributions and densities.

3.1.5.3. Database Integration

After having explored and pre-processed the toponymic data, the data is integrated into a database as outlined in this section.

The modified copy of the toponymic data is stored in a database in form of a table consisting of columns that contain toponymic information and variables necessary to carry out the research tasks. The columns could be named as suggested in section 3.1.1, e.g., *longitude*, *latitude*, *versionI*, *versionII*, *versionIII*, and *fc*. Data type, length or precision, and possibly scale must be defined for each column. The scale is defined if the spatial information is stored as geographic coordinates (longitude and latitude). Scale is the number of digits to the right of the decimal point in a number and is set to 6 for geographic coordinates while length or precision is set to 9.

This step enables the researcher to easily create new columns, perform queries, and to assign new values to the data.

3.2. Development of a Toponym Classification System

This section covers the development of a toponym classification system regarding linguistic origin and feature type and deals with the following research question:

> **RQ1** *How can toponymic data be classified for GIS-based analysis and visualisation by linguistic origin, generic meaning, and geographical feature type?*

The development of a classification system for toponyms is an essential part of the research process and represents one of the research questions. Classifying the toponymic data facilitates the creation of toponymic subsets and the application of GIS-based methods in the following steps. This section is divided into four parts which cover (3.2.1) how the linguistic origins of toponyms are identified by their morphemes, (3.2.2) how the feature types of the original toponymic data are aggregated based on their morphemes' generic meanings, (3.2.3) how new classes are created and (3.2.4) how the data is updated by querying the database. The first two steps focus on gaining a theoretical understanding whereas the last steps follow up on the steps outlined in section 3.1.5 and deal practically with the toponymic dataset.

3.2.1. Identification of Linguistic Origin by Using Morphemes

The first part of the toponym classification deals with the identification of the Indigenous linguistic origin of toponyms. For this purpose, this research suggests using morphemes and its allomorphs belonging to Indigenous languages that are influential in the toponymy of the selected study area.

Given the lack of etymological information in large toponymic datasets, according to Yeginbayeva et al. (2016), this research suggests using Indigenous-derived morphemes to assign linguistic origin to the toponymic data. This methodology is suggested for extensive toponymy research. For intensive toponymy research, investigating the linguistic origin of toponyms individually would be more appropriate. Using the methodology suggested in this research, the toponymic data is queried by morphemes and allomorphs considering their position in the toponym – occurring e.g., as prefixes, or suffixes. The names of the morpheme, and language group are assigned to the respective queried data records. The values must be well-defined for each group.

Querying toponyms for selected morphemes in order to assign their linguistic origins comes with difficulties, including misclassification. Therefore, the selected morphemes are as *unique* as possible to avoid confusions with morphemes or words from other languages. *Uniqueness*, however, is hard to define. In this research, a morpheme's *uniqueness* depends on the decision of the researcher based on her or his knowledge of the study-area specific toponymy obtained from literature analysis in section 3.1.2.

3.2.2. Feature Type Aggregation Based on Generic Meaning

The second part relates to the aggregation of feature types into broader feature type groups. This step is optional. If the feature types or classes in the original dataset are generalized enough according to the researcher and the research purposes, the original feature types can be used for the next steps.

In this research, feature types are aggregated into broader and more generalized feature type groups based on the generic meaning of the morphemes and their associated feature types as outlined in section 3.1.2. For example, if one of the selected morphemes has the generic meaning *mountain*, all feature types in the original dataset associated with mountains could be aggregated into one feature class. In this example, toponyms denominating features, such as hills, mountain ranges, or volcanos, could be aggregated into a feature class called *orographic features*. The values (feature type group names) assigned to the new feature type groups are well-defined and the feature types from the original data that are aggregated in this step are documented to improve traceability.

The result is new feature type groups including feature types that associate with the generic meaning of the morphemes selected during literature analysis.

3.2.3. Class Creation

This section deals with the creation of new classes in the modified table, which has already been integrated into the database according to section 3.1.5.2.

The table is altered by adding new columns representing new classes that will contain information on the Indigenous-derived morpheme(s) included in the toponym, the Indigenous language(s) the toponyms are assigned to, and (optionally) the new feature type group the toponym belongs to. The new columns are created storing the value 0 as default; they will be updated in the next step by data queries (see section 3.2.4). The data type of the new columns can be e.g., string, such as variable character (*varchar*), or numerical, such as integer (*int*). Each value (each group name) that will be stored in these columns (classes) is encoded and defined in an external table.

The following list outlines the suggested class names and the information they contain[4]:

- *morphemeI:* morpheme group name based on morpheme and its allomorph(s) encountered in the toponym
- *morphemeII:* second morpheme group name based on morpheme and its allomorph(s) encountered in the toponym (updated in case there is already a value stored in *morphemeI*)
- *langI:* language group name based on linguistic origin of morpheme and its allomorph(s) encountered in the toponym
- *langII:* second language group name based on linguistic origin of morpheme and its allomorph(s) encountered in the toponym (updated in case there is already a value stored in *langI*)
- *ft* (optional): new feature type group name based on feature type of original data aggregated into new feature class as explained in section 3.2.2

If one toponym contains two morphemes deriving from the same language, the columns *langII* and *langI* have the same value. Furthermore, *langI* (or *langII*) only receives a value, if *morphemeI* (or *morphemeII*) also

[4] The class names will be used throughout this chapter for demonstration purposes. However, they are a suggestion.

contains a value. This is because the values in classes *morphemeI* and *morphemeII* determine the values in classes *langI* and *langII*.

3.2.4. Querying and Updating Database

The last section of developing a toponym classification system covers how the toponymic data is updated by database queries.

Queries are performed to update the values of the new classes created in the step before. For example, all records (toponyms) carrying a certain morpheme (or allomorph) are queried and given a defined value (group name) in the morpheme class *morphemeI*. As the language the toponym derives from is based on the morpheme encountered in the toponym, the language group name is also assigned in the column *langI* accordingly. In case there is already a value stored in these classes from previous querying (value ≠ 0), the values are assigned to the classes *langII* and *morphemeII* instead. This avoids the overwriting of values from previous data updating. For the feature type aggregation, all feature types that form a new – more generalized – feature type group are queried and assigned a new value in the column *ft* to represent the new group.

After having updated and classified the toponymic data by assigning values to the new classes using data queries, the table is exported and imported into a GIS.

3.3. GIS-Based Analysis and Visualisation

This subchapter deals with the methodology used to analyse and visualise the toponymic data and additional datasets using GIS software.

3.3.1. GIS Preparation

In this section, the toponymic data and additional datasets are imported into GIS, a suitable map projection is selected, and toponym subsets are created based on their morpheme group.

The classified toponymic data is imported into GIS. To display the toponyms in GIS, the spatial information stored in the toponymic data needs to be accessed. Furthermore, the additional datasets of historical and present Indigenous language distributions, environmental variables, and boundary layers are imported into GIS. A map projection is selected. This research suggests using an equal-area projection to maintain area measure. For analysis, toponyms are selected by their value stored in the morpheme class; a subset is created from the selection. This step is done for all morpheme groups. The result is several toponymic subsets, each containing toponyms of the same morpheme group. The data can now be analysed and visualised in GIS software.

3.3.2. Overview Maps

This section outlines the creation of maps that provide an overview of a toponymic distribution within the whole study area.

Toponyms of each morpheme group are presented in an overview map as dot symbols, along with additional textual information about the linguistic origin, the generic meaning, and the toponym locations of the morpheme group. Boundary layers provide spatial reference. Furthermore, secondary content, such as:

- title (of the morpheme group),
- possibly subtitles (language group),
- legend,
- scale and scale bar,
- north arrow,
- data sources,
- author(s),
- and publishing date

are added to the overview map. Clusters identified in the next step (see section 3.3.3) are represented by a circle that encloses the toponyms constituting an area of interest (AoI).

3.3.3. Cluster Analysis

In this research, point cluster analysis aims to identify areas of interest (AoIs) within the study area and to define whether a distribution is dense or dispersed.

A point cluster analysis that allows the user to predefine parameters is applied to the toponymic subset of a morpheme group. The selected tool should enable the user to select the two parameters: (1) the number of minimum points, and (2) the search distance which considers the maximum distance between points. The parameters define how clusters are formed and are selected in an iterative process. After this process, the final parameters are set to define the clusters. The toponyms within an identified cluster constitute an AoI; in other words, the extent of the AoI is determined by the spatial location of the toponyms and their extent as a clustered distribution. Again, a new subset of toponyms forming a cluster is created. The subset is analysed in further steps on a scale that is larger compared to the scale of the original subset. Either kernel density estimation along with the isopleth mapping method or toponym aggregation using the hexagon mapping method is applied in further analysis. Which approach is used depends on the final selected parameters in the cluster analysis and the number of toponyms within a cluster (see section 5.2.2).

To sum up, AoIs are identified using point cluster analysis. Subsets of toponyms included in an AoI are created; these are analysed in further steps.

3.3.4. Georeferencing a Historical Map and Dot Mapping

This research uses georeferencing as a method to prepare the analysis of toponym distributions and their relationships with a variable obtained from a historical map. In the case of this research, the variable describes historical language areas represented as polygons. Section 2.2.4.2 creates the methodological framework for the georeferencing process.

The georeferencing process requires a base map as reference including geographical features that are depicted on the historical map. For a small-scale historical map, the features might represent localities or mountain peaks. Usually, GIS software comes along with several base map options. For the georeferencing, control points are selected based on the suggestions formulated in section 2.2.4.2. Afterwards, the accuracy of the selected control points needs to be assessed. The result from this step is a georeferenced raster of the historical map. This raster is digitized; in the case of this research, the language areas are redrawn and saved as a new polygon layer. This layer is necessary for the analysis of the spatial relationship of toponyms with historical language areas covered in section 3.3.5.1.

The georeferenced map is used as a base map. Toponyms are visualised on top of this base map using the dot mapping method. The georeferenced raster gives additional context by providing information on a variable extracted from the historical map. The result is a historical map featuring toponyms visualised as dots. Boundaries are displayed as reference.

3.3.5. Analysis and Quantification of Spatial Relationships

This section deals with the suggested methodology used to analyse and quantify spatial relationships of toponymic distributions, language distributions, and the geographic environment. The approach evaluates the spatial relationship based on intersection and distance to the corresponding features. The section is divided into two subsections, each dealing with one of the following research questions:

RQ2: *How can the spatial relationship of the Indigenous language spoken in a region with the toponyms deriving from this language be analysed?*

RQ3: *How can the spatial relationship of the generic meaning of a toponym with its geographic environment be analysed?*

3.3.5.1. Toponyms and Language Distribution

This subsection deals with the analysis and quantification of the spatial relationship of the toponyms' linguistic origins (based on morphemes) and geographic language distributions. This approach calculates the number of toponyms located inside or within a defined distance to historical and current language areas.

A subset is selected that represents an AoI of a toponymic distribution identified in section 3.3.3 by point cluster analysis. The distribution contains toponyms of the same morpheme group and consequently of the same language group. Furthermore, this analysis requires information on the geographic distribution of the language that the toponyms derive from (according to the language group). For this purpose, the digitized language layer obtained from the georeferenced historical map is activated. Only polygons are selected that represent the area of the language from which the toponyms derive. All toponyms located inside or within a defined distance to the historical language area are counted and divided by the number of all toponyms within the subset. In this step, a search radius must be defined. The search radius is chosen by the researcher and depends on, e.g., scale of the AoI, and accuracy of the georeferenced historical map. The higher the accuracy, the lower the uncertainty of the placement of language borders – theoretically spoken. However, in real life, language borders do not end abruptly. For this reason, a search radius is considered in any way. This ensures that toponyms, which are located relatively nearby the language borders, are also included in the calculation.

In addition to the information of historical language areas, a layer representing the current language distribution is activated. The approach to quantify the spatial relationship of toponyms with present-day language distributions depends on the symbol type that represents the distribution. In case the variable is stored as polygons, the same approach is applied as for the historical language layer: all toponyms located inside or within a defined distance to the current language areas are counted. To compare the result to all toponyms of a subset, it is divided by the overall toponym count.

The result of the calculations is expressed as a ratio and percentage of toponyms located inside or within a defined distance to (1) the historical language areas and to (2) the present-day language areas and the total toponym count. The results from the quantitative analysis are stored in a table.

For visual analysis, the language layers and the toponyms are symbolised and represented in a map. The current language layer is visualised by filling the language polygons with a color. The historical language layer extracted from the historical map is visualised as a polygon with a colored and dashed outline, but without fill color. Therefore, the historical language border appears to be a line that is placed on top of the current language areas. Also, the toponyms are depicted as dots on top of the other layers using the dot mapping method. Boundaries are added to provide spatial reference.

3.3.5.2. Toponyms and Geographic Environment

This subsection deals with the analysis and quantification of the spatial relationship of the toponyms' generic meanings (based on morphemes) with the geographic environment. This approach calculates the number of toponyms located inside an area where the environmental variable is present (for an *areal-based approach*) or within a defined distance to the environmental variable (for a *line-or-point-based approach*).

Besides a toponymic subset, this approach requires a dataset on an environmental variable that represents the generic meaning of the morpheme present in the toponyms (collected in section 3.1.2). The number of toponyms located inside the areas, where this environmental variable is present, is calculated. This approach is used for a dataset storing the environmental variable as polygons. This approach is termed *areal-based* in this research. An exemplary dataset would be elevation data that stores height information within different polygons. For point or line features, the number of toponyms located within a defined distance to the features is calculated. For example, the relationship of toponyms carrying the meaning *river* in their morphemes is analysed using a dataset of the hydrographic network. This approach is denominated as *line-or-point-based* in this research. In both approaches, the calculated number of toponyms is divided by the overall toponym count in the subset. The results are stored in a table.

To visually analyse the relationship of toponyms of a certain meaning with the geographic environment, the environmental data is symbolised and mapped. The toponyms are placed on top of the environmental layer using the dot mapping method. Depending on the complexity of the environmental variable, the dots might not contain a filling – or at least a transparent filling. Also, boundaries are added for spatial reference.

To sum up, the approach of analysing and quantifying spatial relationships of toponyms with language distribution and toponyms with geographic environment consists of calculating the number of toponyms located inside or within a defined distance to the variable that is assumed to be spatially related to the toponym locations. The result is then divided by the overall count of toponyms in a subset. To visualise the spatial relationship, the toponyms are mapped as dots on top of either language borders or an environmental variable.

3.3.6. Analysis of Toponymic Distributions Differentiated by Feature Type Group

This subsection covers the application of kernel density estimation, isopleth mapping, and hexagon mapping to toponymic distributions of different feature type groups and refers to the fourth research question.

> **RQ4**: *How can distributions of toponyms of different feature type groups be analysed and compared in relation to the overall toponym subset?*

This research applies KDE and the isopleth mapping method to clustered toponymic distributions to visualise toponym density estimates. The hexagon mapping method is applied to dispersed toponymic distributions with a wide spatial extent; it aggregates toponym points within hexagons to visual toponym densities. By visualising toponym density estimates, and toponym densities respectively, both approaches aim to avoid overlapping toponym dots, which appear using the dot mapping method.

3.3.6.1. Kernel Density Estimation and Isopleth Mapping

The approach to visualise toponym density estimates applies kernel density estimation on toponymic subsets and uses isopleths to visualise its results.

First, KDE is applied to a toponymic subset that represents a clustered distribution of toponyms of the same morpheme group identified in section 3.3.3 by point cluster analysis. In this approach, the result is classified using a classification method and selecting a number of classes. Method and class number are selected by the researcher; the findings in the literature review presented in section 2.2.4.6 can be helpful for the selection. Afterwards, the toponymic subset is divided by feature type group. KDE is applied to toponyms of the most and second most common feature type groups. Both KDE results are classified using the same class number and ranges as for the KDE result of the whole toponymic subset. To analyse and compare different toponym density estimates and spatial extent among the two most common feature type groups visually, the two KDE results are visualised using the isopleth mapping method. The symbology is the same for both KDE results that are mapped in two separate maps. For further visual comparison, the historical language layer extracted from the historical map are visualised on top as a polygon with a dashed and colored outline, but without fill color. Also, visualising the boundary layer provides spatial context.

The final maps visualise the KDE results for toponyms of the two most common feature type groups symbolised as isopleths whose areal values are taken from the KDE result of the whole toponymic dataset.

3.3.6.2. Hexagon Mapping

This research aggregates toponyms into hexagonal grid cells to visualise toponym densities. In contrast to the approach outlined in section 3.3.6.1, this approach is applied to a dispersed toponymic distribution.

A hexagonal grid is generated for the AoI with an appropriate hexagon size; the size of the hexagons is selected by the researcher according to the scale of the AoI. For a small-scale analysis, the size of the hexagons might be larger than for a large-scale analysis. The toponym count within each hexagon is calculated. The hexagons are classified according to their toponym count; for this step, the findings outlined in section 2.2.4.6. are helpful to find a suitable classification method and number of classes. Then, the toponymic subset is divided by the two most common feature type groups. The toponym counts of the most and second most common feature type groups are calculated for each hexagon. The toponym counts for the two feature type groups are symbolised using the same class breaks and class numbers as obtained from the classification of the whole toponymic subset. The results are visualised in two separate maps. Spatial context is given by visualising the historical language layer and boundaries on top of the hexagonal grid.

The final maps visualise the toponym density within regular shaped areas; they visualise the toponym counts of the two most common feature type groups of a toponymic subset using the hexagon mapping method.

3.3.7. Final Visualisation – Map Sheets

This section covers the creation of the final visual representations in the form of a map sheet. The map sheet combines the results from the application of analysis and visualisation methods in a common layout. Each map sheet represents the results for an AoI, which is defined as explained in section 3.3.3 and contains toponyms of the same morpheme group.

First, a suitable layout is required to present the findings in five different maps. The layout depends on scale and extent of the AoI. A layout suggestion is to arrange the maps in form of a grid. Each grid cell contains one of the five maps. The analysis and visualisation methods, the content presented in these maps, and the related research question are specified in Table 1.

Map No.	Primary map content	Methods	Research Question
Map 1	Georeferenced historical map and toponymic distribution	Georeferencing a historical map, dot mapping	
Map 2	Toponymic, historical and current language distributions	Visualising language distribution, digitizing variable extracted from georeferenced map, dot mapping	RQ2
Map 3	Map of geographic environment (visualises one environmental variable) and toponymic distribution	Visualising environmental data, dot mapping	RQ3
Map 4	Toponym density estimate of most common feature type group and historical language border or Toponym density of most common feature type group and historical language border	Digitizing variable extracted from georeferenced map and KDE and isopleth mapping method or hexagon mapping and toponym count aggregation	RQ4
Map 5	Toponym density estimate of second most common feature type group and historical language border or Toponym density of second most common feature type group and historical language border	Digitizing variable extracted from georeferenced map and KDE and isopleth mapping method or hexagon mapping and toponym count aggregation	RQ4

Table 1 Methods used to create five different maps and their primary map content

Each map may include a boundary layer(s) representing municipality, state, or country borders for spatial reference. Furthermore, the tables created in sections 3.3.5.1 and 3.3.5.2 are presented along with the map sheets to add the results from the quantitative analysis of the spatial relationships visualised in maps 2 and 3. Also, for each map within a map sheet the scale is adjusted so that the AoI fits into the corresponding map grid cell. Secondary content, such as:

- title (of the morpheme group and cluster number),
- legends,
- scale and scale bar,
- north arrow,
- data sources,
- author(s),
- and publishing date

are added to the map sheets.

4. Toponyms in Mexico

This chapter outlines the application of the suggested methodology to toponymic data of Mexico using additional datasets of language distributions, environmental variables, and boundary layers. The fourth chapter is divided into four sections. The first section describes the research preparation. The second part addresses the methodology that was used to classify the toponymic data. The third section outlines the application of GIS-based analysis and visualisation methods to the classified toponymic data of Mexico and other datasets. The last section introduces the expert interviews.

4.1. Research Preparation

The first section of this chapter deals with the research preparation and is divided into (1) an overview of the study area, (2) datasets, (3) software, (4) literature analysis, and (5) the preparation and exploration of the toponymic data.

4.1.1. Study Area – Overview

This section gives an overview of the study area to which the methodology is applied.

The study area includes Mexico. It focusses on Indigenous-derived toponym groups that form AoIs within Mexico. Mexican toponymy has been constituted during different historical eras and is linguistically influenced by several Indigenous languages, the Spanish language, and various combinations (Israel et al., 2016; León-Portilla, 2010). Figure 28 shows the linguistic origins of Mexican federal state names. 15 out of 32 Mexican state names derive from Indigenous languages. Of those 15 states, nine derive from Nahuatl, three from the Mayan language family, two from Purépecha which is a language isolate, and one state name derives from Cora/Nayeri (Heraldo de México, 2019; León-Portilla, 2010). Some state names are commemorative, their origins are rather explained by historical naming processes than language itself. The map (Figure 28) gives an overview of the study area and shows the diverse linguistic origins of Mexico's federal state names. The toponymy of the study area Mexico is further outlined in section 4.1.4.

Figure 28 Toponymy of Mexican Federal State Names by Language Origin

4.1.2. Datasets

This section addresses the datasets used in this research.

The toponymic information for Mexico is based on the Geographic Names Database that contains official standard names approved by the United States Board on Geographic Names and is maintained by the National Geospatial-Intelligence Agency (NGA) (GNS, 2022). The dataset contains 484,013 records (toponyms) and 36 columns, which include information about toponym location (longitude and latitude), toponym standard spellings and spelling variants, feature class, and designation code (e.g., populated place, stream, mountain).

Historical language areas were extracted from the ethnographic map by Manuel Orozco y Berra, which was created in the 19th century with the intention to reflect the Indigenous reality at the time. The map is undated but associated with the book *Geography of Languages and Ethnographic Letter of Mexico, preceded by an essay on the classification of the same languages and notes on the immigration of the tribes* (translated from Spanish) originally published in 1864 (Orozco y Berra, 2003). The map is digitally available in the Mexican Digital Library (BDMx) (Biblioteca Digital Mexicana A.C., n.d.).

The first and second most widely spoken Indigenous languages on municipality level were downloaded from the geoportal of the Mexican Commission for the Knowledge and Use of Biodiversity (CONABIO, 2004). From the same data, municipality borders were extracted.

The federal state borders of Mexico originate from NaturalEarth as large-scale vector data (1:10 m) for the level *Admin 1 – States, Provinces*.

The hydrographic network data was taken from the Mexican National Institute of Statistics and Geography (INEGI, 2006). Also, the data on elevation was taken from (INEGI, 1990).

4.1.3. Software

This section is about software and programming languages used in this research.

For the preparation of the data, this research used R programming language and RStudio Desktop as integrated development environment. Structured Query Language (SQL) was used for data queries. The toponymic data was queried within the relational database management system PostgreSQL. The web-based graphical user interface pgAdmin was used to interact with the Postgres database. This research used the desktop GIS application ArcGIS Pro developed by Esri for GIS-based analysis and visualisation.

4.1.4. Literature Analysis

In this section, a study-area-specific analysis of literature focusses on answering the questions outlined in section 3.1.2. For the literature analysis of this research, the work conducted by León-Portilla (2010), Lefebvre & Paredes Martínez (2017), and Estremo et al. (n.d.) has proven to be of exceptional value.

The study area encompasses a large area with a complex toponymy (León-Portilla, 2010). According to Tichelaar (2002), Mexico has 293 native languages; it is one of eleven countries with more than 200 different native languages. This is reflected by a toponymy with diverse linguistic origins, particularly of Indigenous and European origin (Israel et al., 2016; León-Portilla, 2010). Some of the influential Indigenous languages conserved in Mexican toponymy that this thesis focusses on are Mixtec, Nahuatl (also known as Mexicano according to the ethnographic map by Orozco y Berra (Biblioteca Digital Mexicana A.C., n.d.), Purépecha, and Maya (León-Portilla, 2010).

Historically, Mexican toponymy is divided into three different layers according to León-Portilla (2010):

1. The oldest layer is dating back to the 2nd and 3rd century BC and includes toponyms originating from more than 100 different Indigenous languages.
2. The second layer characterizes toponyms that are superimposed by Nahuatl names as a consequence of the establishment and domination of groups speaking this language. In some cases, Nahuatl toponyms are translations of pre-existing place names in other languages. Therefore, Nahuatl toponymy is also described as expansive.
3. The third layer, which is more complex, was created with the arrival of people from Spain. This layer is characterized by the following changes: alteration of Indigenous toponyms and their written representation due to different pronunciations by Spanish people, combinations of Indigenous and Spanish words in toponyms, and the introduction of toponymy from regions in Spain with several linguistic and historical origins.

Around the time of Mexican independence, the main elements of Mexican toponymy according to León-Portilla (2010) were:

1. Spanish toponyms with multiple linguistic origins (e.g., Roman, Germanic-Gothic, Arabic, etc.)
2. Nahuatl toponyms standing individually or preceded by a Spanish word (often the name of a Saint)
3. Toponyms from different Indigenous languages (other than Nahuatl) standing individually or preceded by a Spanish word

The Indigenous language Mixtec particularly influences the toponymy of the federal state of Oaxaca in Southern Mexico (León-Portilla, 2010). The language holds a large set of toponyms; however, some of the Mixtec toponyms were superimposed by Nahuatl words which are in several cases a translation from Mixtec (León-Portilla, 2010). For example, the Mixtec toponym *Yucu dzaa* which means *bird hill* (translated from Spanish: *colina del pájaro*) was transformed into the Nahuatl toponym *Tototepec* which means *bird hill* or *bird mount* (translated from Spanish: *monte o colina del pájaro*) (León-Portilla, 2010, p.175).

As stated before, Nahuatl language has a large impact on Mexican toponymy. Around the time of colonialization, Nahuatl toponymy extended from Central Mexico towards the South (León-Portilla, 2010). However, Nahuatl-derived toponyms exist all over the country today (León-Portilla, 2010).

Languages of the Maya language family have largely influenced the toponymy of the Yucatán Peninsula (León-Portilla, 2010). In contrast to other Indigenous toponymies, Mayan toponymy has undergone little alteration with exception of some changes in colonial and modern times (León-Portilla, 2010).

Purépecha has largely influenced the toponymy in the state of Michoacán; however, the language was also spoken in the states of Guerrero, Querétaro, and Guanajuato (León-Portilla, 2010).

Morphemes that derive from the Indigenous languages Mixtec, Nahuatl, Maya, and Purépecha were selected during literature analysis. Focus laid on morphemes with a generic meaning that refers to the geographic environment (one exception is the Purépechan morpheme -cuaro which means *abundance*, see Table 2). They were compiled along with their allomorphs, literature references, feature types associated with the generic meaning, and environmental variables that are represented by a dataset. This compilation is shown in Table 2 and was created according to Table 2. The morphemes that were analysed were considered *unique*. This means that they are not easily confused with morphemes from other languages (see section 3.2.1). Unfortunately, this research came across many morphemes found in toponyms originating from Indigenous languages that are not unique. Main reason is that they are too short, consisting of only two or three letters. One example is the suffix *-chi* dominant in the language of the Rarámuri (or Tarahumara) living in the state Chihuahua in Northern Mexico (León-Portilla, 2010, p.176). However, this morpheme is also dominant in Spanish-derived toponymy as part of the word *chico, -a* which means *small* in Spanish. Also, the locative suffixes *-to*, *-o*, and *-ento* in Purépecha do conflict with many Spanish toponyms.

language	**morpheme group name**	**morphemes/ allomorphs + position**	**generic meaning**	**reference**	**associated feature type**	**environ-mental variable**
nahuatl	tepetl	-tepetl	-tepetl = mountain, mountain range, hill (translated from monte, sierra, cerro in Spanish)	Lefebvre & Paredes Martínez (2017, p. 409)	orographic features	elevation
		-tepec	-c = locative suffix			
		-tepeque	-que = locative suffix			
	zoqui	-zoqui-	zoquitl = mud (translated from lodo in Spanish)	Lefebvre & Paredes Martínez (2017, p.271)	hydrographic features	hydrography
	chichil	-chichil-	chichiltic = red, colorado (translated from colorado, rojo	Lefebvre & Paredes Martínez (2017, p.450-451)	places with reddish soils, such as iron	

			in Spanish)		oxide or ocher solis; reddish waters	
mixtec	yucu	-yucu-	yucu = hill, mount, heap, mountain range, weed that spreads and throws leaves (translated from cerro, monte, montón, sierra, "yerba que se extiende y echa hojas" in Spanish)	Alvarado (1962), as cited in Lefebvre & Paredes Martínez (2017, p.450-451)	mountain, hill, ridge	elevation
	yuta	-yuta-, -yute-	yuta, yute = river (translated from río in Spanish)	Alvarado (1962), as cited in Lefebvre & Paredes Martínez, (2017, p.407); Erickson de Hollenbach (2013), as cited in Lefebvre & Paredes Martínez (2017, p.407)	river	hydrography
maya	chen	-chen	chen = spring (translated from pozo in Spanish)	(Estremo, 2022)	spring	hydrography
purépecha	cuaro	-cuaro	cuaro = locative suffix with meaning of abundance (translated from abundanica in Spanish)	(Estremo, 2022)		

Table 2 Compilation of selected Indigenous-derived morphemes

4.1.5. Toponymic Data Preparation and Exploration

This section deals with the exploration and preparation of the toponymic data according to section 3.1.5.

The toponymic dataset for Mexico was downloaded from the GEONet Names Server (GNS) as ZIP file named MX.zip (MX represents the isocode for Mexico). It contains one TXT file including all toponyms of the study area (484,013 records); eight TXT files including toponyms disaggregated by their feature class (fc): A = administrative (country, state, region...), H = hydrographic (stream, lake...), L = localities (parks, area...), P = populated places (city, village...), R = transportation (road, railroad...), S = spot (building, farm), T = hypsographic (mountain, hill, rock...), and V = vegetation (forest, heath); and a disclaimer file. In this research the TXT file including all toponyms was used. It was first converted into a CSV file by replacing tab spaces by semicolons and opened in RStudio where columns not essential for this research were removed to keep the data volume and computation time low. Furthermore, NA-values were replaced by value 0. The remaining eight columns are shown in Table 3. The columns contain the following information:

- **lat** = latitude specifying north-south position of a point on the Earth's surface
- **long** = longitude specifying east-west position of a point on the Earth's surface
- **fc** = feature class (8 different feature classes)
- **dsg** = designation code as a subgroup of feature class (317 different designation codes each belonging to one feature class)
- **adm1** = Numerical code representing federal state; for Mexico from 1 to 32 (some of them have the value 0)
- **sort_name_ro** = Reading order; upper-cased; without spaces, hyphens, or diacritics
- **full_name_nd_ro** = Reading order, without diacritics
- **sort_name_rg** = Reversed reading order; upper-cased; without hyphens, or diacritics

	LAT	LONG	FC	DSG	ADM1	SORT_NAME_RO	FULL_NAME_ND_RO	SORT_NAME_RG
1	18.18996	-87.86025	H	CHNM	0	BOCABACALARCHICO	Boca Bacalar Chico	BACALARCHICO BOCA
2	17.82891	-90.77720	H	LK	0	LAGUNAELTORO	Laguna El Toro	ELTORO LAGUNA
3	17.90926	-88.85418	H	STM	0	RIOAZUL	Rio Azul	RIOAZUL
4	15.80458	-91.91660	H	STM	0	RIOAZUL	Rio Azul	AZUL RIO
5	17.76018	-91.42439	H	STM	0	RIOSANPEDRO	Rio San Pedro	SANPEDRO RIO
6	16.47866	-90.54490	H	STM	0	RIOSALINAS	Rio Salinas	RIOSALINAS
7	14.53147	-92.22813	H	STM	0	RIOSUCHIATE	Rio Suchiate	RIOSUCHIATE
8	17.81535	-90.98770	S	BP	12	VERTICECAMPECHE	Vertice Campeche	VERTICECAMPECHE
9	17.88333	-88.86667	H	LK	0	LAGUNAAZUL	Laguna Azul	AZUL LAGUNA
10	17.85000	-90.38333	H	STM	0	RIOPAIXBAN	Rio Paixban	PAIXBAN RIO
11	15.55681	-92.07969	H	STMI	0	RIOHOJABLANCA	Rio Hoja Blanca	RIOHOJABLANCA
12	15.13240	-92.10886	T	VLC	0	VOLCANTACANA	Volcan Tacana	VOLCANTACANA
13	16.10478	-91.44752	H	STM	0	RIOIXQUISIS	Rio Ixquisis	RIOIXQUISIS
14	32.98333	-110.78333	H	STM	0	RIODESANPEDRO	Rio de San Pedro	SANPEDRO RIODE
15	31.33333	-111.08333	T	MTS	0	SIERRADELPAJARITO	Sierra del Pajarito	PAJARITO SIERRADEL
16	32.15000	-113.73333	T	MTS	0	SIERRADELTULE	Sierra del Tule	TULE SIERRADEL
17	15.58333	-92.16667	H	STM	0	RIOTAPIZALA	Rio Tapizala	TAPIZALA RIO

Table 3 First data records of the pre-processed toponymic dataset for Mexico in RStudio

To explore the pre-processed toponymic dataset, it is imported into GIS, along with census data on a municipality level, and state and country boundaries. Aim is to explore the overall distribution and relative toponym densities in the study area. Kernel density estimation was applied to the toponymic data; the result was visualised as isopleths (Figure 52). Also, the toponym count within a municipality was divided by the population within this area (extracted from census data); the result was visualised as a choropleth map (Figure 53).

After the pre-processing and visual exploration of the toponymic data in GIS, the modified data was saved as CSV file and a database was created with the pgAdmin client in the database management system PostgreSQL. Using the exported CSV file, a new table was created in the database defining its properties. In this step the permissions to access the file were changed. Using a Microsoft Windows operating system this was accomplished by opening the file properties and changing permissions to everyone. Also, a connection between RStudio and the PostgreSQL database was established using the R package *RPostgreSQL*. This way, queries could be directly performed in RStudio using SQL language. This was useful to explore the queried output directly in RStudio by applying R functions.

4.2. Toponym Classification

After having pre-processed the toponymic data, the baseline for toponym classification and value assignment was created as outlined in this section. This section addresses the creation of new classes in the database, assignment of language and morpheme groups to the toponymic data, and the aggregation of feature types. This section refers closely to the output generated from literature analysis outlined in section 4.1.4. Furthermore, this section concerns the classification of the toponymic dataset essential to answer the first research question (**RQ1**).

4.2.1. Class Creation

To prepare the toponym classification, a copy of the table (which was created in the previous section) was generated in the database; it was altered by adding five more columns which contain the value 0 as explained in section 3.2. The properties of the new table including column names, data types, length/precision, and scale are shown in Figure 29. The columns received names according to section 3.2.3 (this section also states the information that the columns would receive in the following steps). Table 4 shows exemplary record entries of the altered table with the five new columns in pgAdmin.

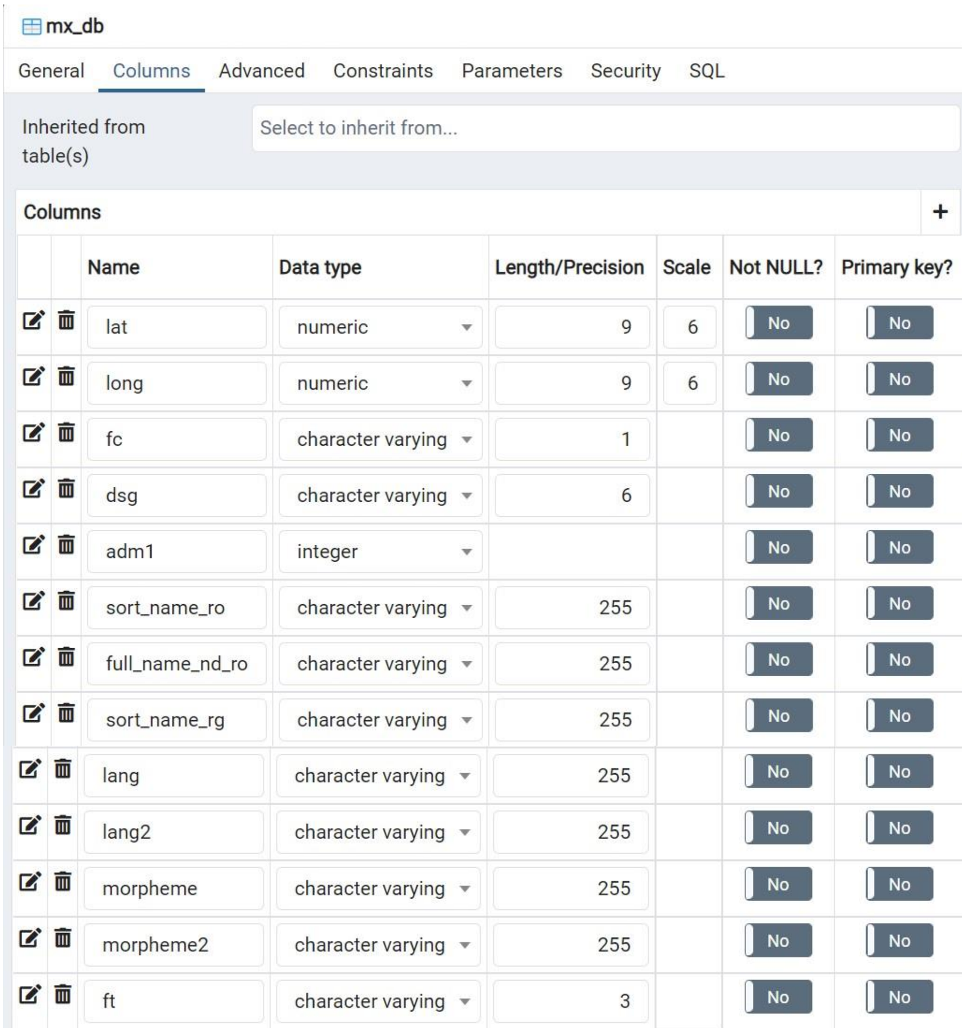

Figure 29 Table properties of the altered table in the database in pgAdmin.

	lat numeric (9,6	long numeric (9,6)	fc charac	dsg characte	adm1 integer	sort_name_ro character varying (255)	full_name_nd_ro character varying (255)	sort_name_rg character varying (255)	lang charact	lang2 characte	morpheme character var	morpheme2 character varying	ft charac
1	17.815348	-90.987698	S	BP	12	VERTICECAMPECHE	Vertice Campeche	VERTICECAMPECHE	0	0	0	0	0
2	31.666667	-106.250000	T	VAL	0	VALLEDEJUAREZ	Valle de Juarez	JUAREZ VALLEDE	0	0	0	0	0
3	31.333333	-109.083333	T	CNYN	0	CANONDEGUADALUPE	Canon de Guadalupe	GUADALUPE CANONDE	0	0	0	0	0
4	31.333333	-108.583333	T	VAL	0	VALLEDELASPLAYAS	Valle de Las Playas	LASPLAYAS VALLEDE	0	0	0	0	0
5	26.550000	-99.166667	S	DAM	0	PRESAFALCON	Presa Falcon	FALCON PRESA	0	0	0	0	0
6	31.500000	-109.583333	T	VAL	0	VALLEDEAGUAPRIETA	Valle de Agua Prieta	AGUAPRIETA VALLEDE	0	0	0	0	0
7	31.333333	-108.833333	T	VAL	0	VALLEDELASANIMAS	Valle de las Animas	ANIMAS VALLEDELAS	0	0	0	0	0
8	32.250000	-114.500000	T	DSRT	0	DESIERTODEALTAR	Desierto de Altar	ALTAR DESIERTODE	0	0	0	0	0
9	28.983333	-103.116667	T	CNYN	0	CANONDESANVICENTE	Canon de San Vicente	SANVICENTE CANONDE	0	0	0	0	0
10	29.294023	-103.943627	T	CNYN	0	CANONELPENASCO	Canon El Penasco	ELPENASCO CANON	0	0	0	0	0
11	16.797423	-90.821185	T	ISL	0	ISLALAPALETA	Isla La Paleta	LAPALETA ISLA	0	0	0	0	0
12	17.815348	-90.987698	S	BP	12	VERTICECAMPECHE	Vertice Campeche	VERTICECAMPECHE	0	0	0	0	0
13	27.699118	-99.746258	S	BDG	0	PUENTEINTERNACIONALSOL...	Puente Internacional Solidarid...	PUENTEINTERNACIONALSOL...	0	0	0	0	0
14	26.051215	-98.034454	S	DAM	0	RETAMAL	Retamal	RETAMAL	0	0	0	0	0
15	16.310547	-90.411101	T	ISL	0	ISLOTEELLINECO	Islote El Lineco	ELLINECO ISLOTE	0	0	0	0	0
16	16.168294	-90.426230	T	ISL	0	ISLOTEDELOSPATOS	Islote de Los Patos	LOSPATOS ISLOTEDE	0	0	0	0	0
17	16.181687	-90.443127	T	ISL	0	ISLOTEELCARIBE	Islote El Caribe	ELCARIBE ISLOTE	0	0	0	0	0

Table 4 Data record entries of the altered table in pgAdmin

4.2.2. Language and Morpheme Group Assignment

After having added new columns to the table, language and morpheme groups were assigned to the toponymic data as outlined in this section.

The toponymic data was queried by meaningful morphemes that were identified and selected during the literature analysis outlined in section 4.1.4 (see Table 2). The morphemes' positions within the toponyms were considered. The corresponding morpheme and language group names were assigned to the queried selection. Figure 30 shows an exemplary query to assign the language group name *nahuatl* and the morpheme group name *tepetl* to all toponyms including the morpheme (in this case, the suffix) *-tepetl* or the allomorphs *-tepec*, or *-tepeque*. The queries were created considering that the morpheme or allomorphs were identified at the end of the toponym, either in column *sort_name_ro* (reading order) or in column *sort_name_rg* (reversed reading order). The query was performed also in the column storing the toponym in reversed reading order due to the occurrence of toponyms that consist of two words. This way, morphemes that occur as suffixes or prefixes were also identified for two-word toponyms. Table 5 depicts an example record (toponym) entry with the updated columns *morphemeI*, *morphemeII*, *langI*, *langII*, and *ft*. In this part it should also be noted that toponyms assigned to a morpheme or language group, may include Spanish words and may consequently be of mixed linguistic origin.

```sql
/* Adding values to columns lang, lang2, morpheme, morpheme2 */
UPDATE mx_db
SET lang2 = 'nahuatl', morpheme2 = 'tepetl'
WHERE morpheme NOT LIKE '0'
    AND
    (SORT_NAME_RO LIKE '%TEPETL'
    OR SORT_NAME_RG LIKE '%TEPETL'
    OR SORT_NAME_RO LIKE '%TEPEC'
    OR SORT_NAME_RG LIKE '%TEPEC'
    OR SORT_NAME_RO LIKE '%TEPEQUE'
    OR SORT_NAME_RG LIKE '%TEPEQUE')
;
UPDATE mx_db
SET lang = 'nahuatl', morpheme = 'tepetl'
WHERE morpheme LIKE '0'
    AND
    (SORT_NAME_RO LIKE '%TEPETL'
    OR SORT_NAME_RG LIKE '%TEPETL'
    OR SORT_NAME_RO LIKE '%TEPEC'
    OR SORT_NAME_RG LIKE '%TEPEC'
    OR SORT_NAME_RO LIKE '%TEPEQUE'
    OR SORT_NAME_RG LIKE '%TEPEQUE')
;
```

Figure 30 Exemplary queries to assign language and morpheme group names to the toponymic data

lat	long	versionI	versionII	versionIII
19.557055	-97.426495	CERROCHICHILTEPEC	Cerro Chichiltepec	CHICHILTEPECCERRO

morphemeI	morphemeII	langI	langII	ft
chichil	tepetl	nahuatl	nahuatl	orographic

Table 5 Exemplary record entry of the table in the database

4.2.3. Feature Type Aggregation

This section deals with the aggregation of feature types.

According to the generic meanings of the selected morphemes, feature types associated with their meaning, and environmental variables that could represent the feature types were identified (see Table 2). To aggregate these feature types and create new generalised feature type groups, the definitions of the feature types in the original data were identified. In this study, the generic meaning of the morphemes refers to hydrographic and orographic features. Consequently, two new feature type groups, *hydrographic* *(hyd)* and *orographic (mts)*, were created. As most toponyms were belonging to the feature class P (*populated places*) and extensive toponym research has focussed on the study of these types of toponyms (e.g., Estremo et al., n.d.; Leibniz-Institut für Länderkunde, 2021), a third feature type group *populated places (pop)* was created.

The new feature type group orographic (*mts*) contains toponyms of the following designation codes stored in the column *dsg*[5]:

ft	fc	dsg	dsg names	dsg description
		HLL	hill	a rounded elevation of limited extent rising above the surrounding land with local relief of less than 300m
		HLLS	hills	rounded elevations of limited extent rising above the surrounding land with local relief of less than 300m
		MESA	mesa(s)	a flat-topped, isolated elevation with steep slopes on all sides, less extensive than a plateau
		MT	mountain	an elevation standing high above the surrounding area with small summit area, steep slopes and local relief of 300m or more
mts	T	MTS	mountains	a mountain range or a group of mountains or high ridges
		PK	peak	a pointed elevation atop a mountain, ridge, or other hypsographic feature
		PKS	peaks	pointed elevations atop a mountain, ridge, or other hypsographic features
		RDGE	ridge(s)	a long narrow elevation with steep sides, and a more or less continuous crest
		SCRP	escarpment	a long line of cliffs or steep slopes separating level surfaces above and below
		SDL	saddle	a broad, open pass crossing a ridge or between hills or mountains
		VLC	volcano	a conical elevation composed of volcanic materials with a crater at the top

Table 6 Description of designation codes in the original toponymic dataset

The new feature type group hydrographic (*hyd*) was considered identical to the feature class H = hydrographic (stream, lake...) in the original data. This is also the case for the feature type group populated places (*pop*) which was considered identical to the feature class P = populated places (city, village...). Instead of creating new feature type groups, the two original feature classes H and P could have been queried. However, for other study areas the new feature type groups might not be identical to the classes in the original dataset. For coherence purposes, this research stored all feature type groups in a new class.

[5] All designation codes are assigned to feature class T in the original dataset.

To update the database and assign the new feature type group names to applicable toponyms within the new class *ft*, the following exemplary queries were performed:

```
 /* Adding values to column fc */
UPDATE mx_db
SET ft = 'mts'
WHERE dsg = 'MT' OR dsg = 'MTS' OR dsg = 'HLL'
    OR dsg ='HLLS' OR dsg ='MESA' OR dsg ='PK'
    OR dsg ='PKS' OR dsg ='VLC' OR dsg ='SDL'
    OR dsg ='RDGE' OR dsg ='SCRP'
;
UPDATE mx_db
SET ft = 'pop'
WHERE fc = 'P'
;
UPDATE mx_db
SET ft = 'hyd'
WHERE fc = 'H'
;
```

Figure 31 Exemplary queries to assign feature type group names to the toponymic data.

After having assigned new class values to the columns in the table, the updated and classified version was exported as CSV file.

4.3. GIS-based Analysis and Cartographic Visualisation

In this section, GIS software was used to analyse and visualise the classified toponymic data for Mexico. The methods used for the analysis and visualisation is outlined in this subchapter.

4.3.1. GIS Preparation

This section outlines the GIS preparation.

The updated toponymic data was imported as a standalone table in CSV format into ArcGIS Pro. To display the toponyms, the tool *Display XY Data* was used where the field *long* (representing longitude) was selected as X Field and *lat* (representing latitude) was selected as Y Field. Then, the additional datasets on current Indigenous language distributions, environmental variables, boundary layers, and the historical ethnographic map (as raster in PNG format) were imported into GIS. The projected national coordinate system *MexicoITRF2008 LCC* was selected. Moreover, toponyms were selected by their value stored in the morpheme group. From the selection, a subset was created storing toponyms of the same morpheme group. This step was performed for all morpheme groups resulting in the number of toponymic subsets equal to the number of morpheme groups. Each subset created a new point layer in GIS.

The data was then analysed and visualised in GIS software.

4.3.2. Overview Maps

This section outlines the creation of overview maps that present the distribution of toponymic subsets within the whole study area along with additional textual information.

An overview map was created for the toponyms of each morpheme group in DIN A4 landscape format. Toponyms were symbolised as red and partly transparent dots. Country and federal state borders with a gray outline and a light-colored filling formed the base map. Clusters that represent an AoI identified in the next step (section 4.3.3) were presented by a red circle enclosing the toponyms that constitute an AoI, along with their AoI number. Additional textual information about the linguistic origin, and the generic meaning retrieved from literature analysis (section 4.1.4) was added. Furthermore, the states and regions where the toponyms are located were stated in the text. For some overview maps, a glyph was added representing the morpheme and its meaning (depending on availability). In the bottom-right corner, the morpheme and allomorphs of the morpheme group were placed on top of a colored background. The color indicates the language from which the toponyms derive. Furthermore, secondary content was added to each overview map (as outlined in section 3.3.2).

Result of this step are several overview maps for toponyms of different morpheme groups. Toponyms in the whole study area were featured along with context information.

4.3.3. AoI Identification

This section covers the identification of AoIs by using point cluster analysis.

The geoprocessing tool *Find Point Clusters* was applied to selected toponymic point layers of morpheme groups in ArcGIS Pro. The clustering method DBSCAN was selected. This method identifies clusters by predefining two parameters: the search distance and minimum features per cluster. To form a cluster, the number of minimum features within the search distance must be identified by DBSCAN. Minimum features per cluster were selected based on the number of toponyms. In any case, the author set the requirement that clusters had to contain at least ten toponyms. The search distance was selected based on the extent of the toponymic distribution. The final parameters were selected after an iterative process of using distinct parameter values. Toponym clusters were detected whose extents were used for the demarcation of the AoIs. KDE and isopleth mapping were applied to dense toponymic distributions (see section 4.3.6.1); hexagon mapping method was applied to widespread and dispersed toponymic clusters (see section 4.3.6.2). Section 5.2.1 describes whether a toponymic distribution was defined as dense or dispersed.

4.3.4. Georeferencing and Digitization of a Historical Ethnographic Map

This research georeferenced an ethnographic historical map that presents language borders in the 19th century (Biblioteca Digital Mexicana A.C., n.d.). The georeferencing process is outlined in this section.

The historical map layer storing a raster image was activated. A topographic basemap integrated in the software was selected. The georeference pane was opened and control points were selected using the basemap as reference. Because the raster concerns a small-scale country map and the country borders did not touch the map frame, locating control points at the corners of the map as suggested in section 2.2.4.2 was not suitable. Additionally, locating control points in Northern Mexico resulted in a high distortion of the original map during the georeferencing process, so control points in Northern Mexico were limited. In total, 28 control points were placed in regions where most toponymic distributions constituted an AoI, which included Central and Southern Mexico. The accuracy was iteratively assessed after deleting control points with a high individual residual error and adding new points to improve the georeferencing result. Table 7 shows the final average discrepancy of the entire set of control points between raster and reference map for various transformations expressed by the total forward root mean square error (RMSE). The georeferencing report can be found in the appendix of section 7.2.1. Although the third order polynomial transformation received the least RMSE and therefore the least discrepancy between control point locations, it transformed the raster image into a highly distorted map (see Figure 32). Therefore, the second order polynomial transformation with an average discrepancy of 22.68 km was the best choice to transform the raster into a georeferenced map (Figure 33). However, if analysis would only include areas close to control points, the third order polynomial-transformed raster could indeed be a better choice. This choice, therefore, depends on the needs of the researcher, AoIs, and research aims.

Transformation	Total forward RMSE [m]	Note
Similarity polynomial	24,348	
1st oder polynomial	23,637	
2nd order polynomial	22,681	Best choice
3rd oder polynomial	18,299	Distorted
Projective	23,502	

Table 7 Total forward RMSE for various transformations of the georeferenced historical map

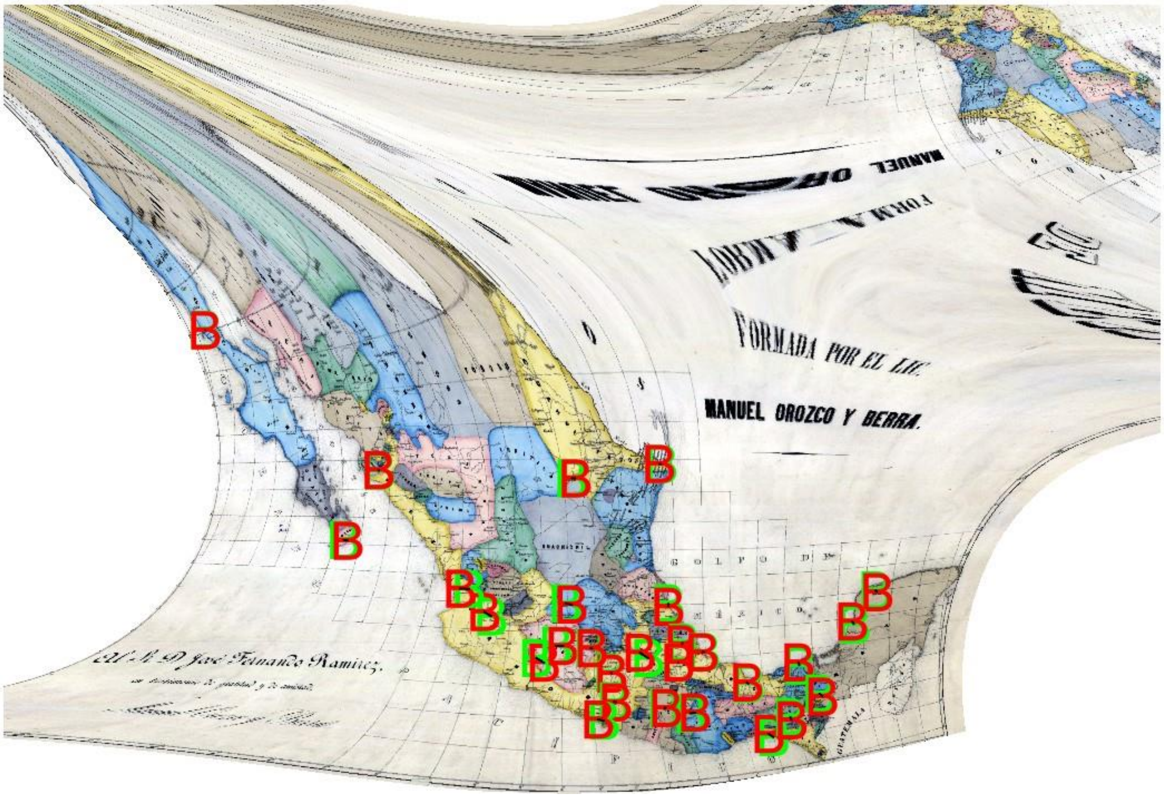

Figure 32 Georeferenced historical map using 3rd order polynomial transformation along with its control points depicted as the red letter B

Figure 33 Georeferenced historical map using 2nd order polynomial transformation along with its control points depicted as the red letter B

The output was exported as a georeferenced raster containing real world coordinates. However, the image resolution of the resulting digital map was rather poor. Due to the unavailability of the whole map in higher resolution, a subset of the historical map with a higher resolution was used. The subset encompasses the region of Central Mexico and was therefore useful for AoIs with toponym distributions in that area. The iterative georeferencing process was conducted a second time for the subset raster image. The accuracy assessment shows similar results (see Table 8); the second order polynomial transformation delivered the most appropriate georeferencing result (Figure 35), whereas the third order polynomial transformation had the least overall RMSE while delivering a highly distorted map (Figure 34). The overall RMSE is lower for the georeferenced subset in comparison to the georeferenced map. The final georeferencing report for the subset can be found in the appendix of section 7.2.2.

Transformation	Total forward RMSE [m]	Note
Similarity polynomial	15,448	
1st oder polynomial	14,373	
2nd order polynomial	11,557	Best choice
3rd oder polynomial	8,219	Distorted
Projective	13,003	

Table 8 Total forward RMSE for various transformations of the georeferenced historical map subset

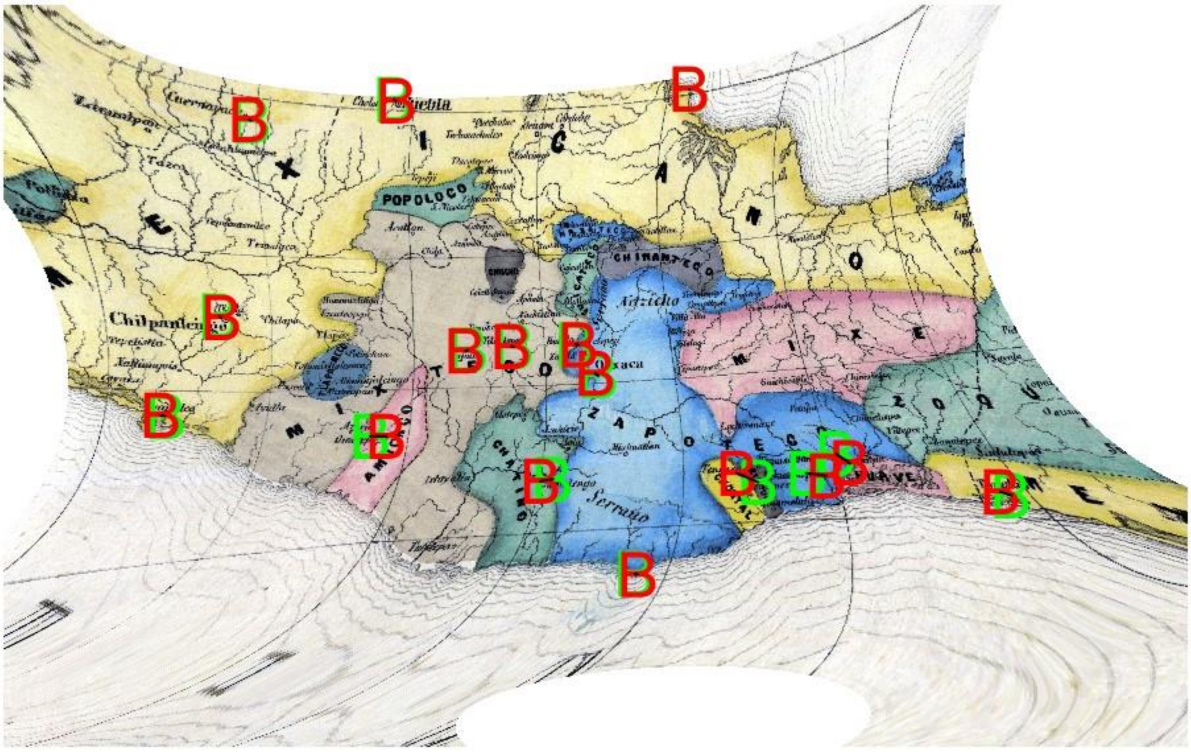

Figure 34 Georeferenced subset of the historical map using 3rd order polynomial transformation along with its control points depicted as the red letter B

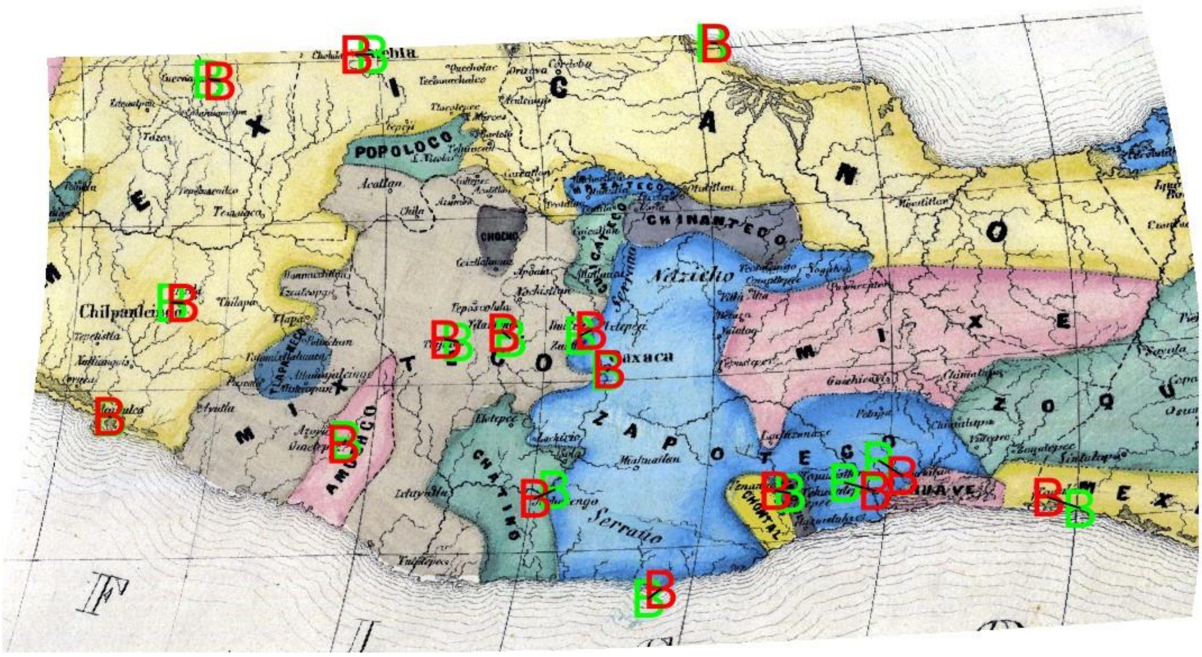

Figure 35 Georeferenced subset of the historical map using 2^{nd} order polynomial transformation along with its control points depicted as the red letter B

For the visualisation, toponyms of a certain AoI were placed as dots on top of the georeferenced map subset (see section 5.1.2; Figure 38, Figure 39 & Figure 40). Furthermore, the areas of the languages from which the morphemes of interest derive from were digitized by redrawing them and saving them as a polygon layer (Figure 36). For this purpose, a new feature class with polygon as geometry type was created in ArcGIS Pro. As the polygon border was masked by the boundary layer in the next steps, areas that touch the coast could reach outside the language area. The masking provided a better visual result as it ends the polygon layer where the masking layer (the boundary layer in this case) starts. This digitization step was necessary for the quantification of spatial relationships in the next steps.

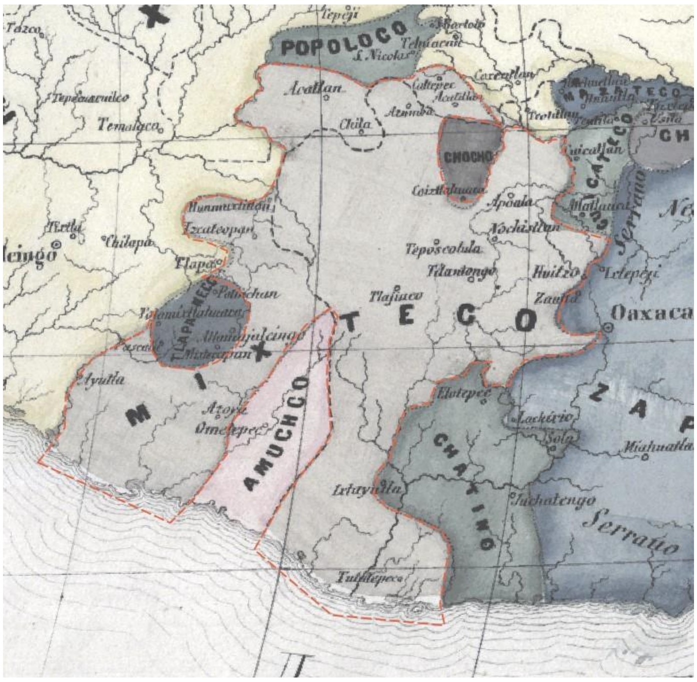

Figure 36 Digitized border of the Mixtec language area

4.3.5. Analysis and Quantification of Spatial Relationships

This section covers the methodology to quantify and visually analyse spatial relationships. The first subsection focusses on the approach to analyse the relationship of the toponyms' linguistic origins with historical and current language distributions. It refers to the second research question (**RQ2**). The second subsection outlines the approach to analyse the relationship of the toponyms' generic meaning with the geographic environment. This subsection refers to the third research question (**RQ3**).

4.3.5.1. Toponyms and Language Distributions

To analyse the relationship of the toponyms' linguistic origins with historical and current language distributions, this research applied an approach that is outlined in this section.

The layers which include the toponyms of a morpheme group within an AoI and the digitized historical border of the language from which the toponyms derive were activated. The toponyms were symbolised as red, medium-transparent dots. The corresponding language area was depicted as an outlined, transparent polygon. Toponyms located inside and within a 10-kilometer distance to the language area were calculated. A buffer of 10 kilometers was selected since language areas do not end abruptly in reality; furthermore, the buffer aims to balance out the inaccuracies owed to georeferencing. The buffer distance was selected subjectively. All calculated toponym counts were divided by the overall toponym count.

Furthermore, the polygon layer representing current language distributions was added to the visualisation; the corresponding language that was either first or second most widely spoken Indigenous language on a municipality-level was symbolised as a choropleth map. Different colors were chosen for the two categories. All other municipalities that did not have the corresponding language either as first or as second most widely spoken Indigenous language were symbolised in white. No outline was given to the polygons. The number of toponyms located inside (intersecting with) and with a 10-kilometer distance to the municipalities (where the language is either the first or second most widely spoken Indigenous language) was calculated. All calculated toponym counts were divided by the overall toponym count.

To calculate the toponym count, the *Select Layer By Location* tool was used in ArcGIS Pro. The toponymic layer (*input layer*) was selected by its intersections with the corresponding language layer (*selecting features*). Therefore, the relationship parameter was set to *Intersect*. In the case for the calculation of toponyms located within a distance to the language layer, the *search distance* was set to 10 kilometers. The results are presented in section 5.1.2: Table 10 shows the toponym counts as ratio and percentage, and the maps in Figure 41 depict the symbolised layers.

4.3.5.2. Toponyms and Geographic Environment

To analyse the relationship of the toponyms' generic meaning with the geographic environment, this research applied an approach that is outlined in this section.

Two layers were activated: the layer of the toponym distribution within an AoI and the layer of the selected environmental variable that is associated with the generic meaning of the toponyms. To describe the *areal-based approach*, this section uses toponyms including the morpheme *yucu* as an example. Since the generic meaning of *yucu* is *hill* or *mount* (among others), an elevation dataset including polygons classified by heights was selected. The height zones were visualised as isopleths filled with a diverging color scheme. The toponyms were placed on top as gray dots. The dots' filling was set to partly transparent, allowing to see through the complex elevation layer. Polygons were selected by their height and the toponyms inside the selected polygons were counted. Their count was divided by the total toponym count. The visualisation result is shown in Figure 43; the calculations are presented in Table 11.

To describe the *line-or-point-based* approach, a toponymic subset of the morpheme group *yuta* is used. Because *yuta* means *river* in Mixtec (see Table 2), hydrographic dataset which represents river streams as polylines was used. The toponyms located within different distances to these polylines were counted. Furthermore, the river streams were symbolised according to their Strahler number (describing the hierarchical order of river streams) using graduated colors and different line widths; the toponyms were placed on top of the river streams as red (almost fully transparent) dots. The visualisation result is shown in Figure 44Figure 43; the calculations are presented in Table 12Table 11.

4.3.6. Analysis of Toponymic Distributions Differentiated by Feature Type Group

This section describes the suggested methodology to analyse toponymic distributions differentiated by their feature type group. The methodology aims to answer the fourth research question (**RQ4**). For this analysis, two approaches were used. KDE and isopleth mapping were applied to densely clustered toponymic distributions. Toponym aggregation using the hexagon mapping method was applied to dispersed toponymic distributions[6]. The following subsections outline these two approaches.

4.3.6.1. Kernel Density Estimation and Isopleth Mapping

This section describes the application of KDE to densely clustered toponymic distributions and the use of isopleth mapping method to visualise the KDE results.

The *Kernel Density* tool was applied to a toponymic subset (point features) in ArcGIS Pro. First, KDE was applied to the whole toponymic subset. The search radius (*bandwidth*), used to estimate the density, was calculated using Silverman's Rule of Thumb (Silverman, 1986) directly in ArcGIS Pro. The output cell size was selected in a way, so that the result held a high spatial resolution while keeping computation time to a minimum. In the case of this research, the cell size was set to 0.001 km^2. All other variables were set to default, including population field (none), output cell values (densities), method (planar), and input barrier features (none). The result was classified using seven different classes and the natural breaks method. Then, the two most common feature type groups – either orographic features (*mts*), hydrographic features (*hyd*), or populated places (*pop*) – were identified. KDE was again applied to toponyms of the two most common feature type groups. Class ranges from the KDE result of the whole toponymic subset were applied to the KDE results of the two different feature type groups. Isopleths filled with graduated colors were used to visualise the KDE result and to symbolise the different toponym densities. The higher the toponym density, the darker the color. The isopleth class with the least density

[6] Section 5.2.1 describes whether a toponymic distribution is defined as dense or dispersed.

was set to transparent; consequently, only six isopleths were visible in the result. Along with the isopleths, the corresponding historical language border was visualised as a dashed line.

The results of this step were two maps visualising the KDE results of the two most common feature type groups of a toponymic subset using the isopleth mapping method (see Figure 48). Using class ranges from the overall KDE result, visualised the two toponym density patterns relative to the overall density of the toponymic subset.

4.3.6.2. Hexagon Mapping

This section describes the aggregation of toponyms inside hexagonal grid cells. The method was applied to dispersed toponymic distributions.

First, a hexagonal map was generated with the *Generate Tessellation* tool in ArcGIS Pro. The example shown in this section used an areal size of 500 kilometers for each hexagon comprising the tessellation. However, the selected hexagon size depends on scale and extent of the distribution. All toponyms of a morpheme class were aggregated within these hexagons using the *Spatial Join* tool in ArcGIS Pro. A new column was added automatically to the attribute table of the output hexagonal grid. The column stored the count of toponyms intersecting with each hexagon. The hexagonal grid cells were classified by toponym count using the natural breaks method, seven classes, and a graduated color scheme. However, the lowest class was visualised transparent. Therefore, only six classes were visible. The higher the toponym count, the darker the color. Then, the toponymic subset was differentiated by its two most common feature type groups (such as in the approach used for the visualisation of isopleths described in section 4.3.6.1). Toponyms of these two feature type groups were again aggregated within the generated hexagons. Class ranges and symbology from the toponym counts of the whole toponymic dataset were used for the visualisation of the aggregated counts for the two feature type groups. The corresponding historical language border was visualised as dashed line and laid on top of the hexagonal layers.

The result of this approach are two maps visualising the toponym count of the two most common feature type groups of a toponymic subset using the hexagon mapping method (see Figure 50). The two toponym count patterns were visualised relative to the overall toponym count of the toponymic subset using the class ranges from the classification result of the whole toponymic subset.

4.3.7. Final Visualisation – Map Sheets

This section covers the creation of a common layout to present the results obtained from GIS-based analysis and visualisation described in the previous sections.

A common layout was created, which consists of five different map frames arranged as a grid in DIN A4 portrait format (Figure 37). The analysis and visualisation results of several toponym groups were depicted in this layout, resulting in so-called map sheets for each toponym group in a specific AoI. The map scales were adjusted, so that the extent of the AoIs would fit each map frame. Furthermore, secondary content, outlined in section 3.3.7, was added to each map frame.

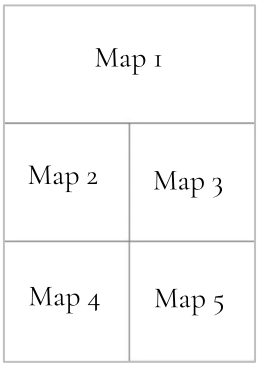

Figure 37 Map sheet layout

Table 9 contains exemplary maps for each map frame in the layout. Also, it includes links to the sections that outline the methodology applied to the study area to create the corresponding map. The research question to which each exemplary map and section refers to, is also stated.

Map No.	**Exemplary maps**	**Section**	**Research Question**
Map 1	Figure 38, Figure 39, and Figure 40	4.3.4	
Map 2	Figure 41 and Figure 42	4.3.5.1	RQ2
Map 3	Figure 43 and Figure 44	4.3.5.2	RQ3
Map 4 and Map 5	Figure 48 and Figure 50	4.3.6	RQ4

Table 9 Context information for each map frame within the map sheet

4.4. Expert Interviews

Two qualitative semi-structured interviews with experts from the fields of linguistics and history were conducted. This chapter introduces the necessity of expert interviews in the context of this research, introduces the interviewees, outlines the meeting setup, and presents the interview guide containing the prepared interview questions.

Since research in toponymy is interdisciplinary, it is important to discuss this research's results together with experts from scientific fields other than cartography or GIS-based research. As Luo et al. (2010, p.3) have formulated it, the study of toponyms are *"an opening hatch to a panoramic view of linguistic and historical patterns of […] settlements and the evolution of […] practices for naming places that are rich reflections of the environment […] and their interaction with other cultures"*. Therefore, expert interviews were conducted in a semi-structured manner to assess the relevance, reliability, and limitations of this research. Two experts from the fields of linguistics and history who are familiar with toponymy research were interviewed. The goal was to gain knowledge of how the results are relevant to researchers from different fields and to answer (some of) the research questions considering the research objectives.

4.4.1. Expert Interviewees

The first expert interview was conducted with linguist Evandro L.T.P. Cunha who has a background in Computer Science (master's degree) and Computational Linguistics (PhD). He works as a professor at the Faculty of Letters of the Federal University of Minas Gerais (FALE/UFMG, Brazil) and does research related to computer science and linguistics. Toponymy is one of his research interests where he has particularly focussed on the toponymy of Brazilian state Minas Gerais. One of his projects is about Indigenous-influenced toponyms in the neighborhoods of Belo Horizonte.

The second interview was conducted with historian Carlos Salvador Paredes Martínez who published the book *The memory of the names: the toponymy in the historical conformation of the territory. From Mesoamerica to Mexico* (translated from Spanish: *La memoria de los nombres: la toponimia en la conformación histórica del territorio. De Mesoamérica a México*) (Lefebvre & Paredes Martínez, 2017) together with Karin Lefebvre. The book is a selection of several presentations given at the colloquium with the same name in 2015. He received his doctorate degree in history from the National Autonomous University of Mexico (UNAM) where he also works at the Center for Research and Higher Studies in Social Anthropology (CIESAS). His line of research comprises ethnohistory, iconography, and the colonial history of Mexico.

4.4.2. Meeting Setup

The meetings with the two experts were conducted online via the video-communication service Google Meet that allows presenting content by screen-sharing. Furthermore, they were recorded using the Windows 10 built-in screen, microphone, and audio recorder. In the following, the meeting setup is outlined:

1. Thanking expert for his time and interest in research and interview.
2. Asking for permission to record the interview.

3. Introducing myself focussing on academic background, research motivation and objectives. Explaining research topic and methodology broadly.
4. Asking expert to introduce himself.
5. Explaining interview setup, which includes:
   (1) Showing results while giving expert the time he needs to explore and understand the content.
   (2) Checking interpretation of expert by asking questions.
   It is stated that throughout the interview, the expert is encouraged to ask comprehension and methodological questions. Also, it is explicitly mentioned that the interview is semi-structured allowing for a dialogue and bidirectional questions.
6. Conducting interview while showing the following results:
   - Figure 55
   - Figure 56
   - Figure 60
   - Figure 61

4.4.3. Interview Guide

An interview guide was created according to Edwards & Holland (2013, p.54 ff.). The guide lists the interview questions:

INTERVIEW GUIDE

1. Do you think that these maps help analysing the relationship of the Indigenous language spoken in a region with the toponyms deriving from this language?
   a. Do you think that these maps could support to indicate a change of language use in a region?
   b. Why do you think is it useful to analyse this relationship of language with toponymy?

2. Do you think that these maps help analyzing the spatial relationship of the meaning of a toponym with its geographic environment?
   a. If yes, to which extent is this relationship analysed?
   b. Why do you think is it useful to analyse the relationship of geographic environment with the generic meaning of toponyms?

3. Do you think these density maps help comparing the quantity and extent of toponyms of different feature types?
   a. Why do you think is it useful to analyse the different feature types of toponyms of a certain morpheme with a meaning?

4. Could you give reasons or explanations from your perspective as a linguist for the spatial pattern of this toponym group by looking at the different maps?

5. Do you think these maps could be reproduced easily with other toponym groups classified based on linguistic origin?
   a. Which adaptations do you think are necessary?
   b. What has to be considered when creating these maps with other toponyms?
   c. Do you think a linguist could easily reproduce them, or do you think a cartographer should be involved in the map creation?

6. Do you think the results are meaningful for linguistic research/research in history?
   a. To which extent?

7. Which limitations do you see in these maps?

8. Have the visualisations become clearer to you now that we have discussed the results and the methodology a bit?

9. Do you have other thoughts?

5. Results and Discussion

This chapter presents the results of this research. It relates the results to the corresponding research objectives, research questions, and hypotheses; embeds them into the context of GIS-based toponymy research; links them to the findings from the expert interviews; and outlines possible directions for future research. Furthermore, this chapter discusses the relevance, reliability, and limitations of this research.

5.1. Research Objectives and Future Work

5.1.1. RO1 – Toponymic Classification System

This section presents the results of the methodology suggested to classify toponymic data according to **RO1**. This section and answers the first research question.

> **RQ1.** *How can toponymic data be classified for GIS-based analysis and visualisation by linguistic origin, generic meaning, and geographical feature type?*

This research developed a methodology to classify a large toponymic dataset by linguistic origin, generic meaning, and geographical feature type. As the dataset did not include any etymological information regarding the toponyms' linguistic origin, morphemes were used to identify the linguistic origin. Morphemes were selected for being known to derive from a particular language, to be influential in the study area's toponymy, and to have a generic meaning referring to the geographic environment. Therefore, the analysis of third-party work on the study area's toponymy was fundamental. The toponymic data was transferred into a database where the data was queried by the selected morphemes. The position of a morpheme within a toponym was considered in the querying process. The queried toponyms were assigned to corresponding morpheme and language groups stored in separate columns. Geographical feature types of the original dataset were aggregated based on the morphemes' generic meanings. New feature type groups were created and assigned to a new column. Feature types that represented the generic meanings of the morphemes were assigned to the corresponding new feature type group. Consequently, this research used a classification system based on morphemes, their generic meaning, and feature type information included in the original dataset to classify toponyms. Table 5 shows an exemplary record entry of the classified toponymic data. The classified toponymic data, originally containing spatial information, was imported into GIS. Analysis and visualisation methods were applied to toponyms differentiated by their morpheme and aggregated feature type group.

The developed methodology has proven useful to classify toponymic data for GIS-based analysis and visualisation by linguistic origin, generic meaning, and geographical feature type; and to answer **RQ1.** The classification scheme based on morphemes and their generic meaning is appropriate for extensive toponymy research. The methodology is useful for large toponymic datasets without etymological information. It enriches the toponymic data with information about its linguistic origin and generic meaning by querying the data for selected and meaningful morphemes. The methodology may be used to explore the toponymy of a large study area and may be preceding to intensive toponymy research.

5.1.2. RO2 – Spatial Relationships of Toponyms and Language Distributions

This section presents the results of the methodology suggested to analyse the spatial relationship of toponyms with language distributions according to **RO2** and **H2**. It answers the second research question along with its sub-question.

> **RQ2** *How can the spatial relationship of the Indigenous language spoken in a region with the toponyms deriving from this language be analysed?*

The methodology suggests the georeferencing of a digital historical map that contains information about historical language distributions. The georeferenced output was visualised along with toponyms of a morpheme group (within an AoI) symbolised as dots. This visualisation gives an overview of the toponymic distribution within an AoI using the georeferenced map as a context-adding basemap. The visualisation (as shown in Figure 38-Figure 40) was featured as map 1 in the map sheet (see Figure 37). Furthermore, the map's attribute information was extracted by digitizing its map features – the historical language areas. The extracted information was used to analyse the spatial relationship of the toponym group with the historical area of the language the toponyms were derived from (according to the language group in the classified data). This relationship was quantitatively evaluated based on intersections and distances of the toponyms with or to the corresponding language polygon (Table 10). The same methodology was used to analyse the relationship of the toponym group with current language distributions. Furthermore, these relationships were visualised by depicting the historical and the current language distribution along with the corresponding toponyms of a morpheme group. The municipalities (of the dataset on current language distributions) where the corresponding language is the first or second most widely spoken Indigenous language were visualised; the historical language area was visualised on top as a transparent (or slightly transparent) polygon with a dashed outline. The language polygon was visualised as slightly transparent for the Nahuatl language area due to ambiguity (Figure 42); a representation as transparent polygon would make it unclear, whether the area outside or inside the dashed outline is the actual language area. This visualisation (as shown in Figure 41 and Figure 42) is featured as map 2 in the map sheet (see Figure 37).

The developed methodology is useful to analyse the spatial relationship of toponyms with the language they derive from and to answer **RQ2**. The hypothesis **H2** that toponyms of a morpheme group are mainly located in areas where the language they derive from was and is spoken is confirmed, at least for the selected morphemes groups *yucu*, *yuta*, and *chichil* within their corresponding AoI number 1.

Linguist Cunha confirms that the suggested methodology is useful for this research purpose, referring to Figure 41: *"It's an interesting visualisation since we see clearly this overlap between the toponyms and the Mixtec region – let's say – as the most widely spoken language [...]. Here it's clear that there is some relationship with the language"* (Cunha, 2022, see transcription of expert interview in section 7.1.1).

However, the results of the analysis of the spatial relationship – for the historical language areas – was highly dependent on the accuracy of the georeferenced map (see appendix 7.2). The more accurate the georeferencing, the more reliable the quantification results. The georeferenced map subset had an overall forward RMSE of 11.56 kilometers. In comparison, the overall forward RMSE of the whole georeferenced map was with 22.68 kilometers higher. This means that the map subset was georeferenced twice as accurate as the whole historical map. Therefore, using the georeferenced map subset for digitization of the language border improved the reliability of the quantification results. However, since language areas do not end abruptly in reality, it is difficult to assess the accuracy and reliability of the language borders already in the original historical map. Validating the spatial relationship quantitatively is therefore linked to imprecision. To balance out these inaccuracies introduced by ambiguous language borders and georeferencing, a buffer of 10 kilometers was selected. However, the result may vary largely depending on the selected buffer distance selected subjectively by the researcher. For other georeferenced data that has clearly defined borders, a buffer radius that equals the overall forward RMSE (obtained from the accuracy assessment) could be selected.

Nevertheless, the georeferenced map on language areas was verified to be an appropriate and relevant source for this research; according to Paredes Martínez, the selected historical ethnographic map by Orozco y Berra *"is fundamental because it is one of the oldest [sources] that gives an overview of languages in Mexico in the 19th century [...] it is a very good source to have a perspective at a given moment in time. It is an important base for consideration."* (Paredes Martínez, 2022, translated from Spanish, see transcription of expert interview in section 7.1.2).

The suggested methodology has proven useful for the analysis of spatial relationships of toponyms with language. However, it requires more knowledge of the processes leading to these toponymic distributions and spatial patterns to explain them. Even though this research does not aim to explain these patterns, the experts had some possible explanations during the conducted interviews. For example, Figure 42 shows that for the morpheme group *chichil* (of Nahuatl-origin), there is more toponyms encountered outside of Nahuatl language areas than for the two Mixtec morpheme groups in Figure 41. In the case of Nahuatl-derived toponyms, one explanation could be explained by the domination of Nahuatl-speaking groups and the superimposition of their toponyms in regions which were also inhabited by other ethnic groups speaking different languages (León-Portilla, 2010; Paredes Martínez, 2022, see section 7.1.2). In the case of the Mexican state Guerrero, *"Nahuas, the Mexica, are conquering this entire region of northern Guerrero, and are imposing their Nahua toponyms – of Nahua origin – despite the fact that there are Chontal and Tlapanec and [people of] other origin, but the Nahuas, being the conquerors, are imposing their toponyms"* (Paredes Martínez, 2022, see section 7.1.2). But language contact due to e.g., commerce, migration, or family relationships, could also be an explanation for toponyms that are located outside their corresponding language areas. This is also, why Cunha considers difficult to answer the following research sub-question (Cunha, 2022, see section 7.1.1):

RQ2.a *Can the suggested methodology be used to indicate a change of language use in a region?*

According to the linguist, to answer this question more information about study-area-specific processes would be required. This is also stated by historian Paredes Martínez who says that it is essential to investigate historical aspects of place naming including the motivation that led to denominate a place in order to explain those toponymic patterns (Paredes Martínez, 2022, see section 7.1.2). However, the historian confirms the relevance of this research sub-question, since language loss is a serious problem leading to the loss of toponyms. He mentions that according to a Chatino colleague of his *"in the towns where Chatino is spoken, people doubt the persistence of toponyms in the Chatino language because people have lost the use of the language. Then the people themselves say, why keep the name of this or that town, of this or that place if we no longer speak the language? If the language is no longer valid and we don't even know what it means?"* (Paredes Martínez, 2022, see section 7.1.2).

Consequently, to answer research sub-question **RQ2.a** a more detailed and interdisciplinary study of toponymic distributions is required. Therefore, **RQ2.a** may be a research question for future work.

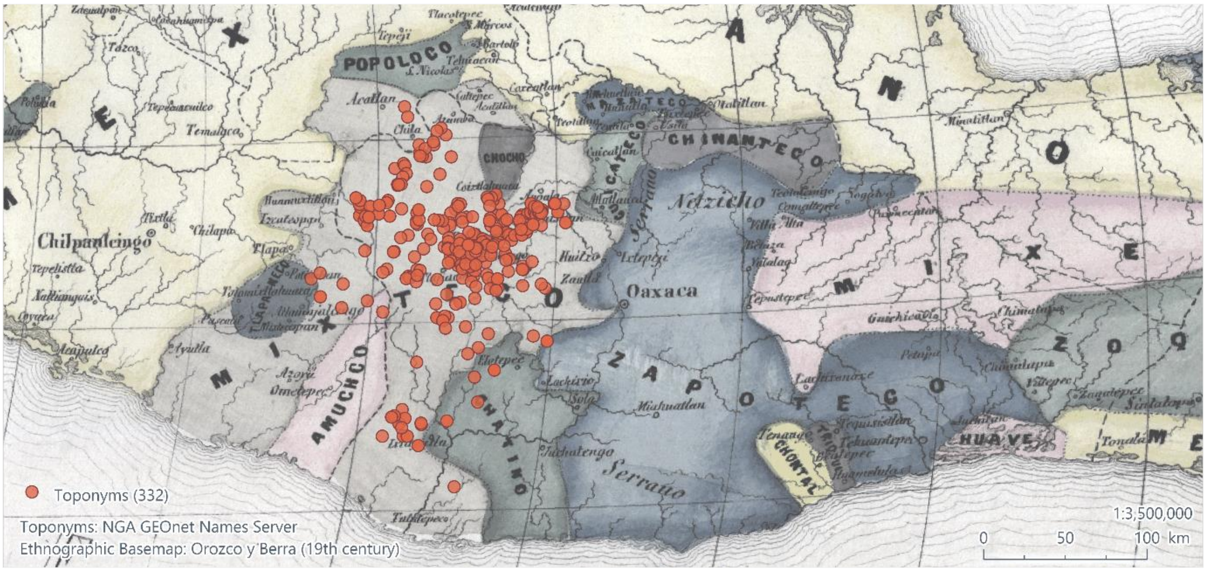

Figure 38 Yucu toponyms visualised as dots placed on top of georeferenced ethnographic map

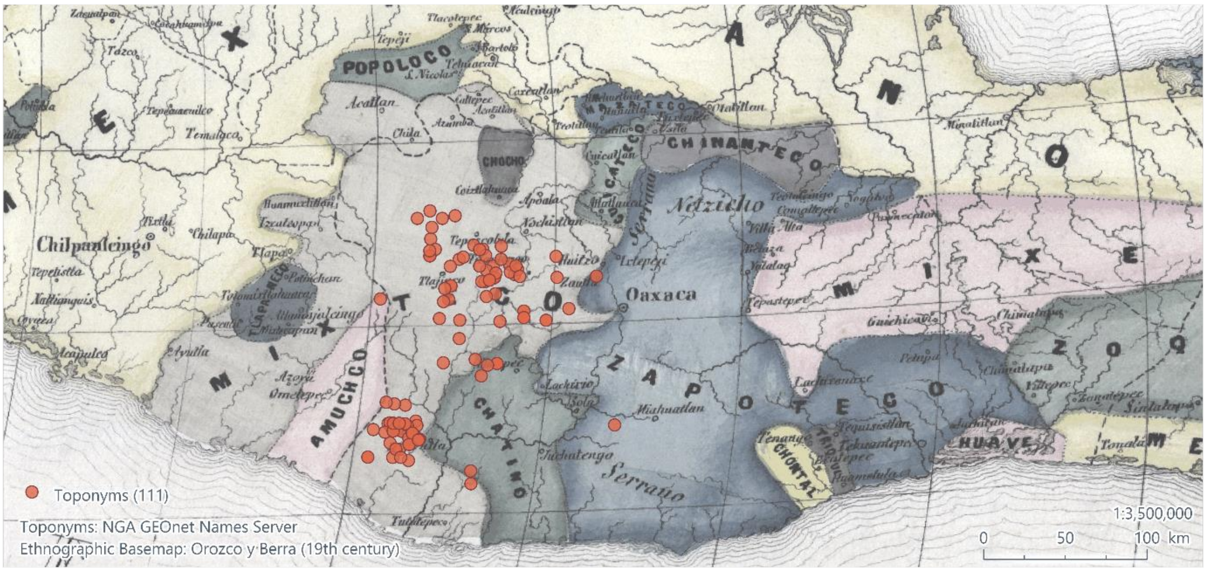

Figure 39 Yuta toponyms visualised as dots placed on top of georeferenced ethnographic map

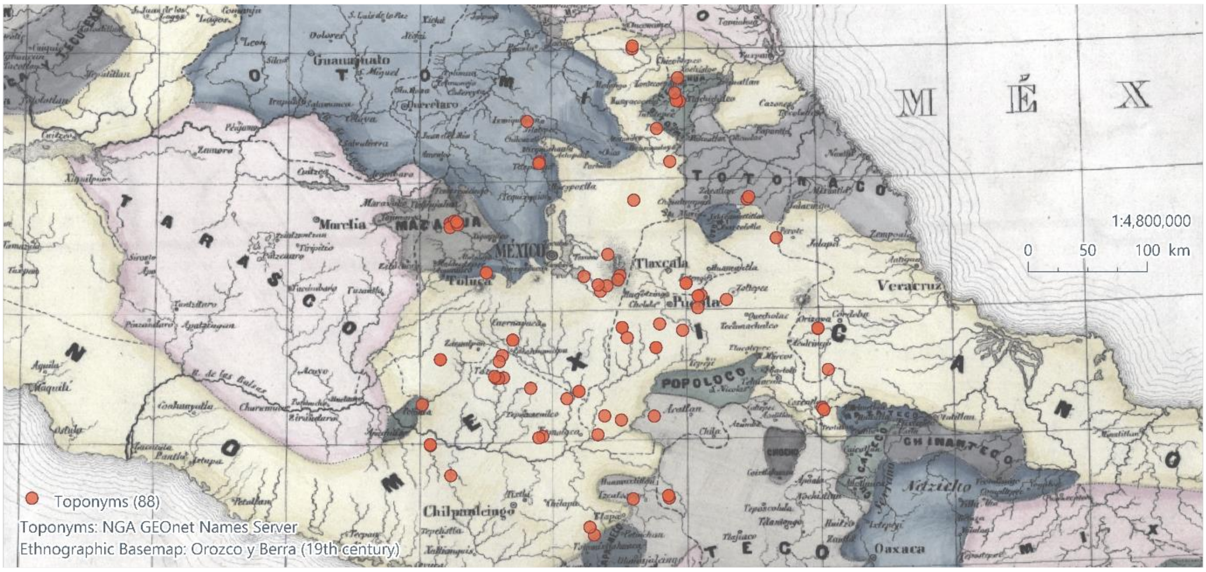

Figure 40 Chichil toponyms visualised as dots placed on top of georeferenced ethnographic map

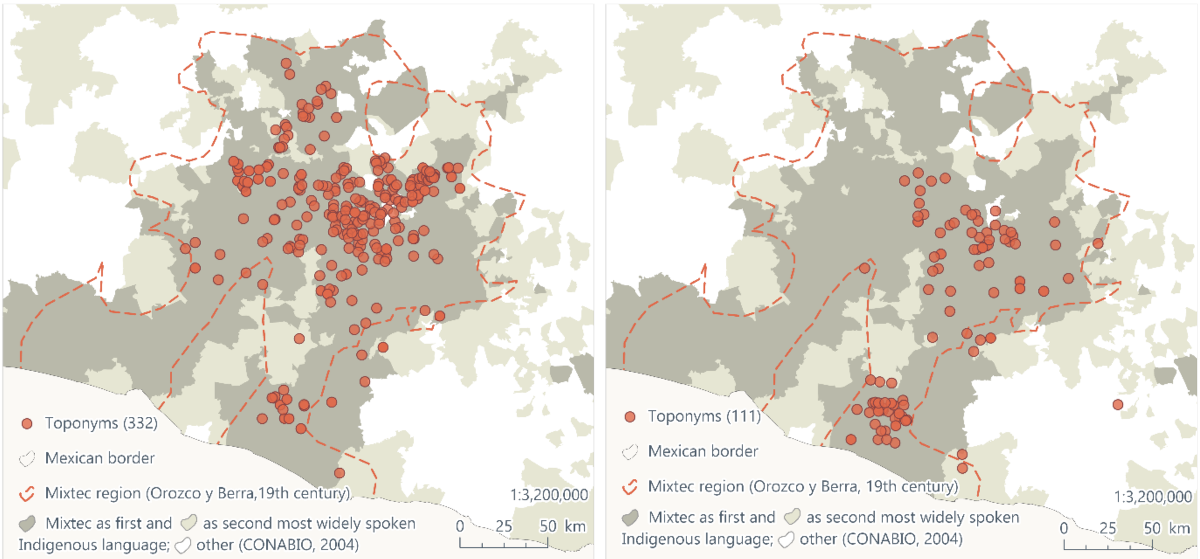

Figure 41 Mixtec language areas and toponymic distribution of yucu morphemes (left) and yuta morphemes (right)

	Language layer				
Toponym subset: yucu – AoI 1	**Historical**		**Present-day**		
	Inside	Within 10-km distance	Inside 1st most widely spoken Indigenous language border	Inside 2nd most widely spoken Indigenous language border	1st + 2nd + within 10-km distance
a: Toponym count	317	327	313	10	332
b: Total toponym count	332				
Ratio (a/b)	0.95	0.98	0.94	0.03	1
Percentage [%]	**95**	**98**	**94**	**3**	**100**

Table 10 Quantification of the spatial relationship of yucu toponyms with language distribution

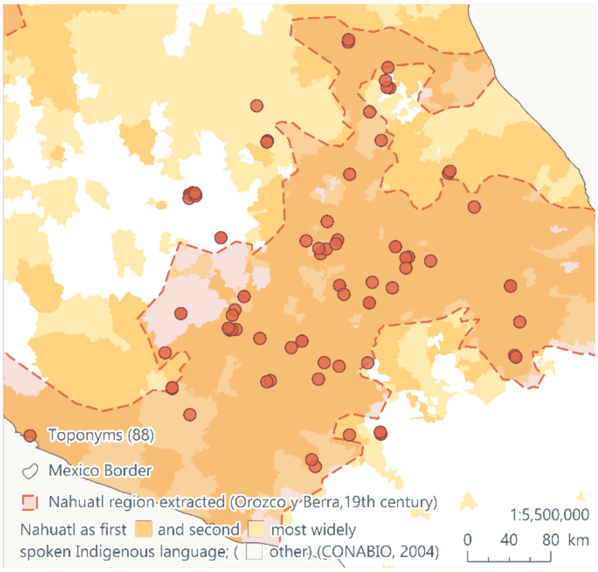

Figure 42 Nahuatl language areas and the toponymic distribution of chichil morphemes

5.1.3. RO3 – Spatial Relationship of Toponyms and Geographic Environment

This section presents the results of the methodology suggested to analyse the spatial relationship of the generic meaning of a toponym with its geographic environment according to **RO3** and **H3.** It answers the third research question.

> **RQ3** *How can the spatial relationship of the generic meaning of a toponym with its geographic environment be analysed?*

To answer this question, this research identified toponyms that carry a selected morpheme. This morpheme has a generic meaning that refers to the geographic environment. The meaning was associated with an environmental variable that could be represented by a dataset. The relationship of toponyms with their geographic environment was evaluated based on intersections with and distance to the environmental variable (such as in the approach to analyse the relationship of toponyms with language). Two different approaches were selected depending on the geometry type of the environmental variable: the *areal-based* approach was used for polygons and the *line-or-point-based* approach was used for polylines (and could be used for point features).

The developed methodology is useful to analyse the spatial relationship of the generic meaning of a toponym with its geographic environment and to answer **RQ3**. The hypothesis **H3** that there is a spatial relationship of the toponym's generic meaning (based on morpheme) with its geographic environment is confirmed, at least for the selected morphemes groups *yucu* and *yuta* within their corresponding AoI number 1. Applying the methodology to these two morpheme groups has shown that there is a high percentage of yucu toponyms intersecting with high elevation zones, and a high percentage of yuta toponyms that is located close to the hydrographic network. The results are discussed in the following paragraphs. The results are shown in the Table 11Table 12, and in the Figure 43 Figure 44, which are featured as map 3 in the map sheet (see Figure 37).

The suggested methodology, including the two different approaches, was applied to two different morpheme groups: (1) *yucu* toponyms whose relationship was analysed by using a polygon layer representing height zones, and (2) *yuta* toponyms whose relationship was analysed using a polyline layer representing the hydrographic network. The intersections of *yucu* toponyms with seven different height zones ranging from zero to more than 3,000 meters were calculated using the *areal-based* approach. It results that 42.5% of *yucu* morphemes are located within the height zone ranging from 2,001 to 2,500 meters, 17.8% were located in the height zone ranging from 2,501 to 3,000 meters (see Figure 43 and Table 11). In future work, all toponyms (disregarding their morpheme group) within the analysed AoI could be included in the analysis. Also, the total distribution of the different height zones in the area could be considered. The inclusion of these two aspects could be useful to find out whether these high percentages of toponyms that are located in high altitudes are only a consequence of: (1) a high-elevation AoI or (2) the fact that toponyms are in general more common in higher altitudes. These tasks including the consideration of these two aspects remain for future investigations.

Furthermore, toponyms of the morpheme group *yuta* were analysed regarding their spatial relationship with the river network using the *line-or-point-based* approach. It results that more than half (55.9%) of the 111 considered toponyms are located within less than 100 meters to the hydrographic network; 92.8% of the toponyms are located within 1,000 meters to the network (see Table 12 and Figure 44). Such as for the areal-based approach, future work could include all toponyms in this investigation; this means that it may be analysed how many toponyms are located within a particular distance, e.g., 100 meters, or 1,000 meters, to the river network. Furthermore, the maximum distance between river streams in the study area could be analysed to find out whether a toponym location close to the river network is consequence of an environment characterized by a large hydrographic network with rivers that are located close to each other. Following on this, future work may investigate the distances of toponyms to river streams of a different order (Strahler number). This would assess whether the relevance or size of a river influences this relationship; in other words, whether toponyms carrying the meaning river in their name are rather located close to small or large river streams.

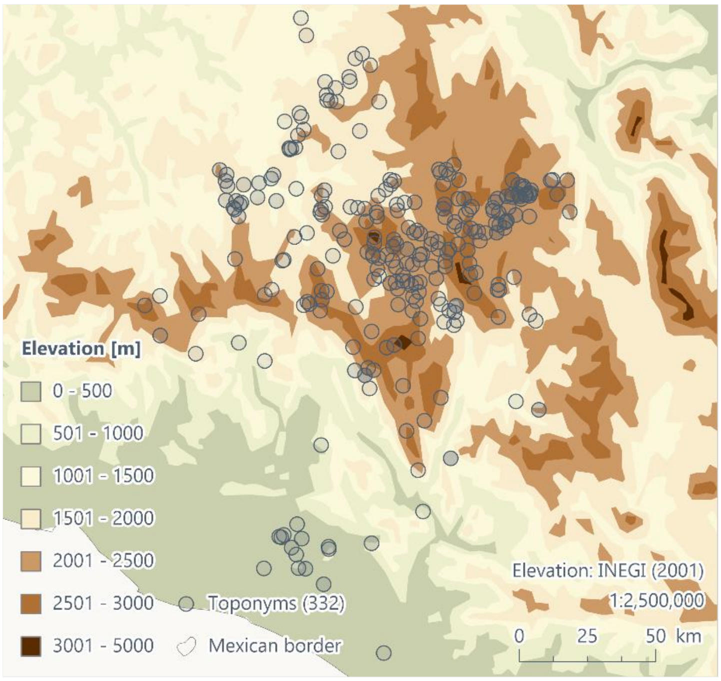

Figure 43 Yucu toponyms placed as dots on top of the elevation layer symbolizing different height zones

	Environmental layer						
Toponym subset: yucu – AoI 1	**Elevation/height zone [m]**						
	≤ 500	501 – 1,000	1,001 – 1,500	1,501 – 2,000	2,001 – 2,500	2,501 – 3,000	≥ 3,001
a: Toponym count within height zone	17	9	19	86	141	59	1
b: Total toponym count	332						
Ratio (a/b)	0.051	0.027	0.057	0.259	0.425	0.178	0.003
Percentage [%]	**5.1**	**2.7**	**5.7**	**25.9**	**42.5**	**17.8**	**0.3**

Table 11 Result of the quantification of the spatial relationship of yucu toponyms with elevation

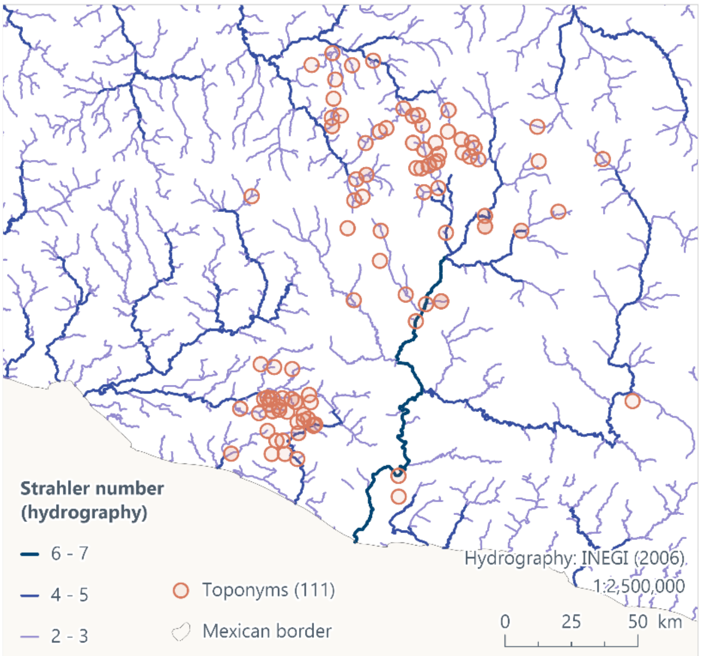

Figure 44 Yuta toponyms placed as dots on top of the hydrographic network symbolised by their Strahler number

	Environmental layer			
Toponym subset: yuta – AoI 1	**Hydrography [m]**			
	≤ 100	101 - 500	501 – 1,000	≥ 1,001
a: Toponym count, within distance	62	29	12	8
b: Total toponym count	111			
Ratio (a/b)	0.559	0.261	0.108	0.072
Percentage [%]	**55.9**	**26.1**	**10.8**	**7.2**

Table 12 Result of the quantification of the spatial relationship of yuta toponyms with hydrography

The methodology and the findings from its application are relevant as this approach sets toponyms into an environmental context; Cunha states that "*when we make this cross comparation between the toponyms and other features in the environment, we gain a lot of information that otherwise [is] hidden when we only see the toponyms without this environmental context*" (Cunha, 2022, see section 7.1.1). Furthermore, he states that *"joining [...] the name information and the environmental information"* is relevant because *"most of the toponyms of Indigenous origin relate to environmental information"* (Cunha, 2022, see section 7.1.1). Also, Paredes Martínez states that *"these comparisons [of toponyms, generic meaning, and geographic environment] are relevant and carrying the toponymic analysis to cartography, I believe that what you are doing is an important step"* (Paredes Martínez, 2022, see section 7.1.2).

Furthermore, the study of the toponym's meaning along with its visualisation in space, provides us with information about the geographic environment. As Paredes Martínez stated it: *"as for someone who [...] is foreign to the place, it [the denomination of a place] means nothing, but [for] the local people it does have a lot of meaning, because it means that water accumulates [at this place] or that [this place] is a very fertile area [...] it is territories that apparently do not tell us anything but that have a lot of meaning locally and this is clearly noticeable in the language"* (Paredes Martínez, 2022, see section 7.1.2).

The following paragraphs present some possible explanations for the spatial patterns. These theories surged during the expert interviews; the statements by the experts are included in the discussion of these theories.

The results show that *yucu* (*mountain* or *hill* in Mixtec) toponyms are mainly concentrated in mountainous regions (Figure 43). However, there is some toponyms in the lowlands, South to these high-concentration areas in the mountains. To explain this agglomeration, analysing the isopleth maps depicting the toponym density estimates of different feature type groups is useful (Figure 46). Cunha expected these toponyms in the lowlands not to be related to orographic features, but to populated places; according to Cunha, these places would carry the morpheme *yucu* to possibly indicate a mountain view (Cunha, 2022, see section 7.1.1). However, in this case, the toponyms' generic meaning would still be related to their environment. Nevertheless, the visualisation shows that the density is slightly higher for toponyms of orographic features in this lowland region (Figure 46). At this point, it is meaningful to closely review the meaning of the morpheme *yucu* again, because *yucu* does not only translate to mountain, but also to *hill* and to *weed that spreads and throws leaves* (Alvarado, 1962, cited by Lefebvre & Paredes Martínez, 2017, p.450). Future GIS-based work could analyse the area in the lowlands more in detail on an even larger scale. To analyse the toponyms' relationship with the meaning *hill* – and consequently with a hilly environment – a dataset on elevation is not sufficient to analyse the toponymic subset's relationship with its environment. A dataset representing isolines with a relatively small interval between the lines could possibly be used for this analysis. The small interval between isolines could be useful to indicate even minor changes of terrain steepness. The result would show whether toponyms are located in an area with minor changes in altitude, or in other words on a *hill* or in a hilly environment. Also, the count of toponyms, which denominate orographic features (214), is more than twice as high as the count of populated places (88). Therefore, the overall density of these features is higher. These counts should also be considered in the analysis.

However, Cunha also mentions the phenomenon of transplanted toponyms, which refers to toponyms that are *transplanted* from one place to another (Cunha, 2022, see appendix 7.1.1). According to the linguist, this phenomenon is usually related to migration and should be considered in analysis. Therefore, a hypothesis to explain the distribution patterns of *yucu* toponyms in relation to elevation data, could be that *"people went from highlands to the coast and then they named the place where they are living at the Coast after the place where they lived in the highlands"* (Cunha, 2022, see section 7.1.1). However, to confirm this hypothesis, additional information on e.g., area-specific migration processes, is necessary (Cunha, 2022, see appendix 7.1.1). Also, Indigenous-derived toponym clusters lying outside of their language area could possibly be the result of toponym transplantation.

The other way around, the suggested methodology could be applied to a study area that is known to be influenced by this process of transplanted toponyms (Cunha, 2022, see section 7.1.1). According to Cunha, it could be analysed whether *"at least some connection with the environment [can be detected] in the naming of these places"* (Cunha, 2022, see section 7.1.1).

The suggested methodology is straightforward and avoids the introduction of uncertainties that come along with complex analysis algorithms. It can be applied to investigate these relationships of further morpheme groups with different environmental variables. However, future work may investigate multivariate approaches to improve the analysis of these relationships.

5.1.4. RO4 – Toponymic Distributions Differentiated by Feature Type Group

This section presents the results of the methodology suggested to analyse distributions of toponyms of different feature type groups and to compare them in relation to the overall toponyms subset according to **RO4** and **H4**. This section answers the fourth research question.

> **RQ4** *How can distributions of toponyms of different feature type groups be analysed and compared in relation to the overall toponym subset?*

To answer this question, this research suggests two approaches to estimate and calculate toponym densities. Which approach was used depended on the distribution pattern of toponyms which was identified using cluster analysis (see section 5.2.2). The overall toponym subset was defined as a subset consisting of toponyms of the same morpheme group within an AoI. The first approach estimates toponym densities by using KDE and visualises the result by using the isopleth mapping method. This approach is applied to densely clustered toponym distributions. For the second approach, the hexagon mapping method is applied to visualise the toponym density of dispersed toponymic distributions. Toponyms were aggregated into hexagonal grid cells visualising their toponym count. In both approaches, the corresponding method was first applied to the overall toponym subset; the result was classified defining the classification method and the number of classes; it was visualised using a graduated color scheme. Afterwards, the subset was differentiated by feature type groups. Again, the corresponding method was applied to the two most common feature type groups. Then, the results were visualised using the same class ranges as defined for the overall toponym subset. Using the same class ranges enables comparison among different distributions of toponyms of different feature type groups; furthermore, this visualises the toponym densities in relation to the overall toponym density. Figure 47a and Figure 49a present the results of both approaches applied to the overall toponym subset. Figure 47b+c, and Figure 49b+c depict the results for the two most common feature type groups using the same class ranges as used for the overall toponym subset. The map visualisation in Figure 48Figure 50 are featured as maps 4 and 5 in the map sheet (see Figure 46).

The developed methodology is useful to compare toponymic subsets of different feature type groups in relation to the overall toponymic subset and to answer **RQ4**. Two approaches were developed for the application to different types of distributions, clustered or dispersed. Furthermore, the creation of subsets based on feature type group revealed that toponyms were not necessarily assigned to the feature type group their generic meaning refers to. Therefore, the hypothesis **H4** stating that toponyms would mainly be assigned to a feature type group their generic meaning refers is rejected. Still, for the morpheme groups *yucu* and *yuta*, the most common feature type group is the one that associates with the generic meaning of the two morphemes: 66 out of 111 *yuta* toponyms are classified as hydrographic features, and 214 out of 332 *yucu* toponyms are classified as orographic features.

Whereas Chloupek (2018) and Zeini et al. (2018) depicted the KDE results along with the toponyms included in the KDE using the dot mapping method (see Figure 20 and Figure 19), this research visualises the KDE result and individual dots using individual map visualisations. A graduated color scheme was also used by Chloupek (2018) see Figure 20, Qian, Kang, & Wang (2016), and Zeini et al. (2018) to visualise the KDE results. Such as in the research by Fuchs (2015) and Zeini et al. (2018) (see Figure 15 and Figure

19), this research used qualitative ordinal class descriptions ranging from low to high density to represent the densities.

According to Cunha, *"it is interesting to compare the density of toponyms with different characteristics [...]"* (Cunha, 2022, see appendix 7.1.1). Furthermore, this methodology indicates regions of high relevance, where toponym distributions of a particular feature type group are denser compared to other toponym groups: Cunha states that *"we don't see these [high densities] in the orographic features [...] this region is more relevant for the populated places than for the orographic feature toponyms"* (Cunha, 2022, see section 7.1.1).

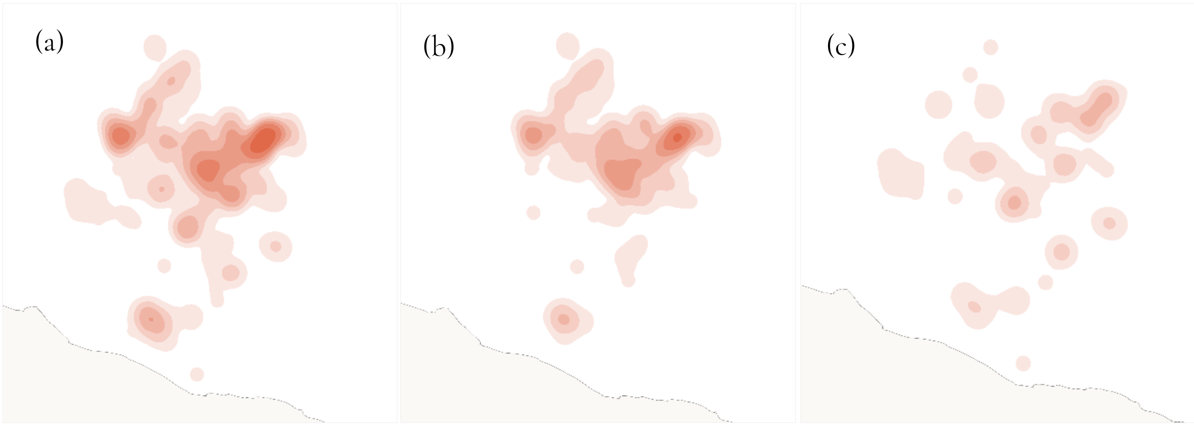

Figure 45 KDE results visualised as isopleths of: (a) the whole yucu toponymic subset, (b) only orographic features, and (c) only populated places

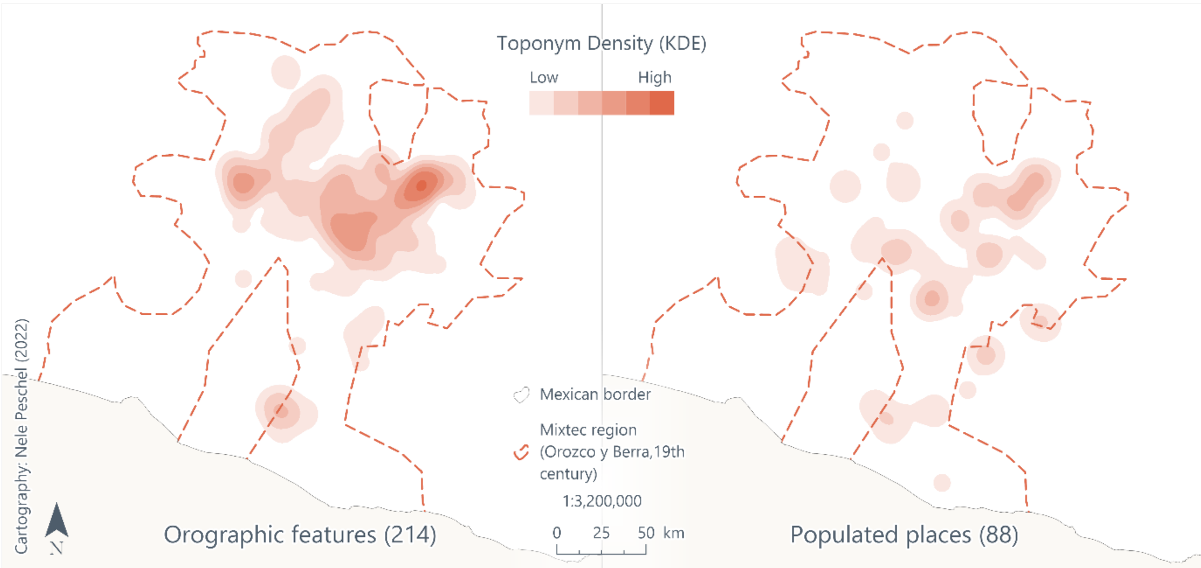

Figure 46 Toponym density estimates of orographic features (left) and populated places (right) of yucu toponyms

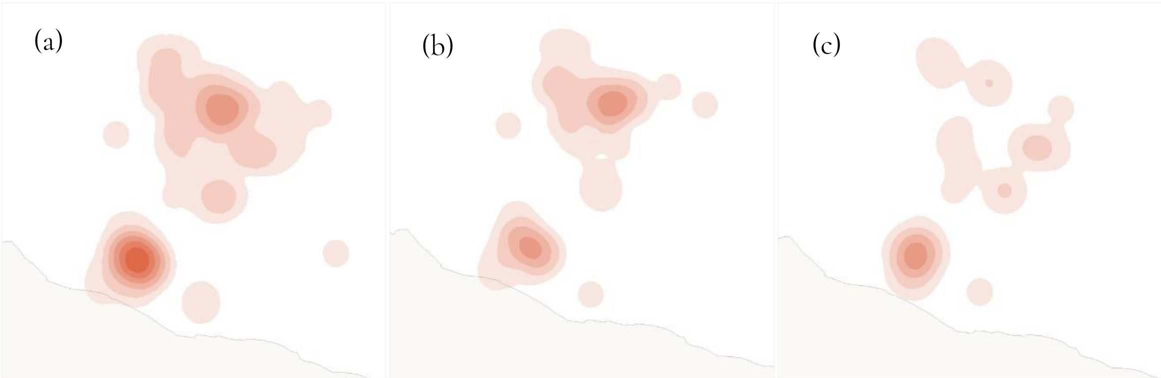

Figure 47 KDE results visualised as isopleths of: (a) the whole yuta toponymic subset, (b) only hydrographic features, and (c) only populated places

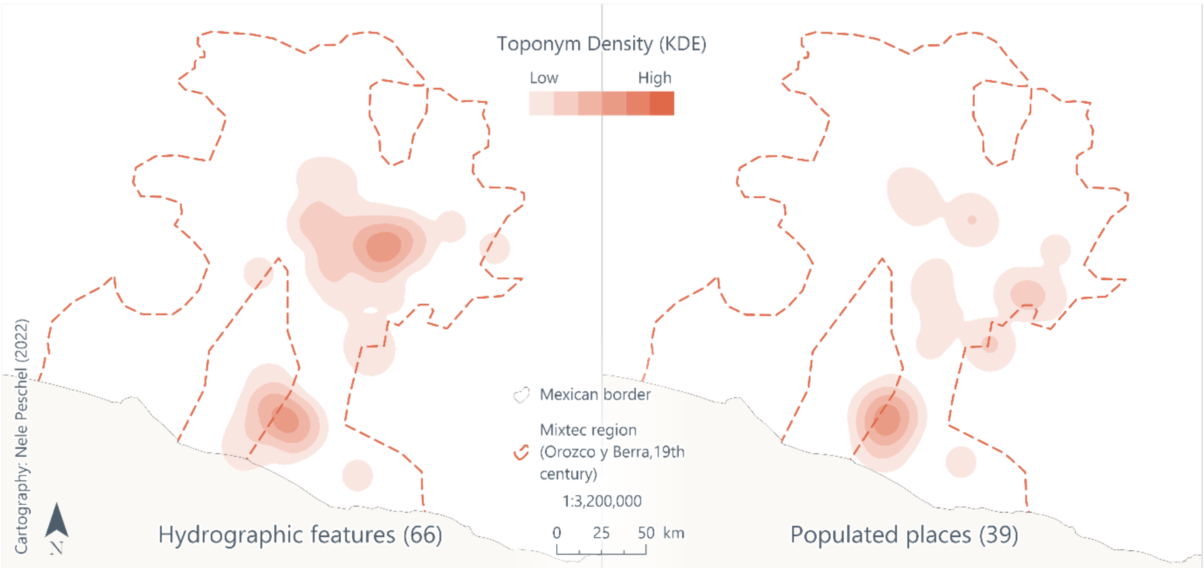

Figure 48 Toponym density estimates of hydrographic features (left) and populated places (right) of yuta toponyms

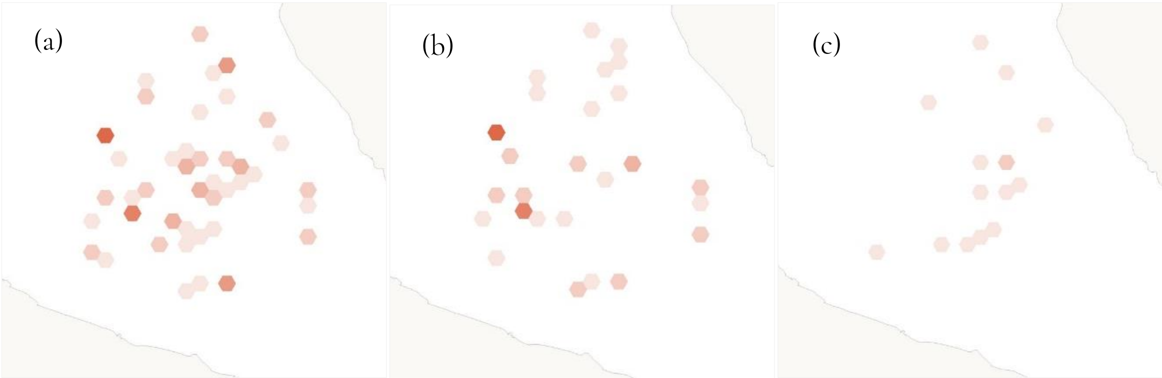

Figure 49 Hexagon mapping method visualises toponym counts for: (a) the whole chichil toponymic subset, (b) only populated places, and (c) only hydrographic features

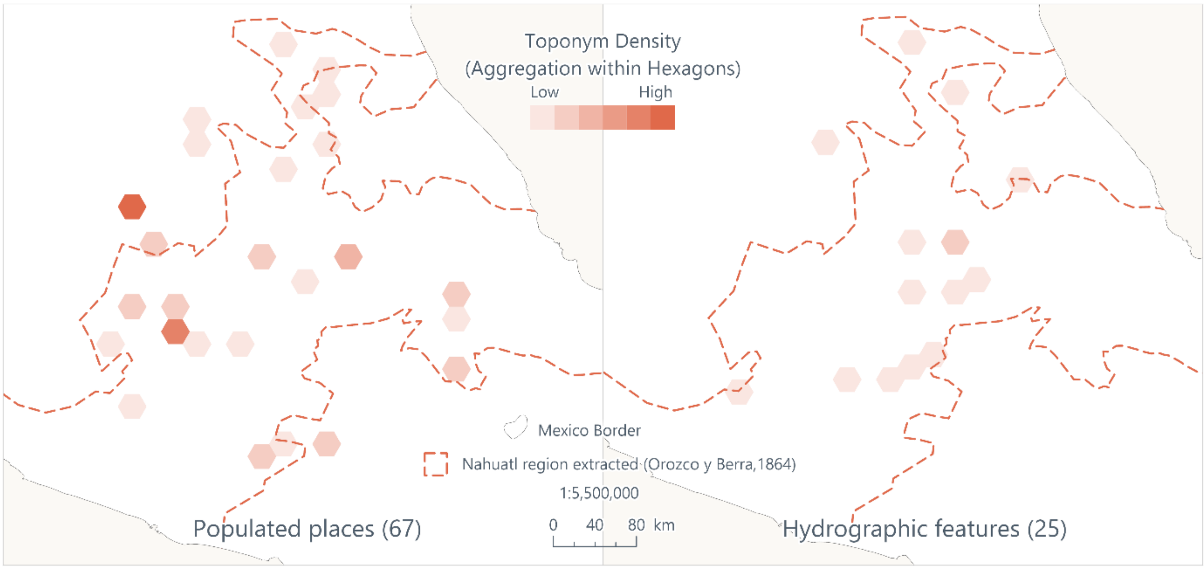

Figure 50 Toponym densities of populated places (left) and hydrographic features (right) of chichil toponyms

The hexagon mapping method to visualise toponym densities is suggested as an alternative to KDE and isopleth mapping. Having applied KDE to the dispersed toponymic distributions of chichil toponyms, the visualisation result of the density estimates represented by isopleths appeared visually busy (Figure 51). Therefore, the hexagon mapping method is suggested to depict toponym counts within regularly shaped sizes. In comparison to the KDE approach, this avoids the representation of toponym density (estimates) between low and high concentrated areas, which is the case for dispersed toponymic distributions.

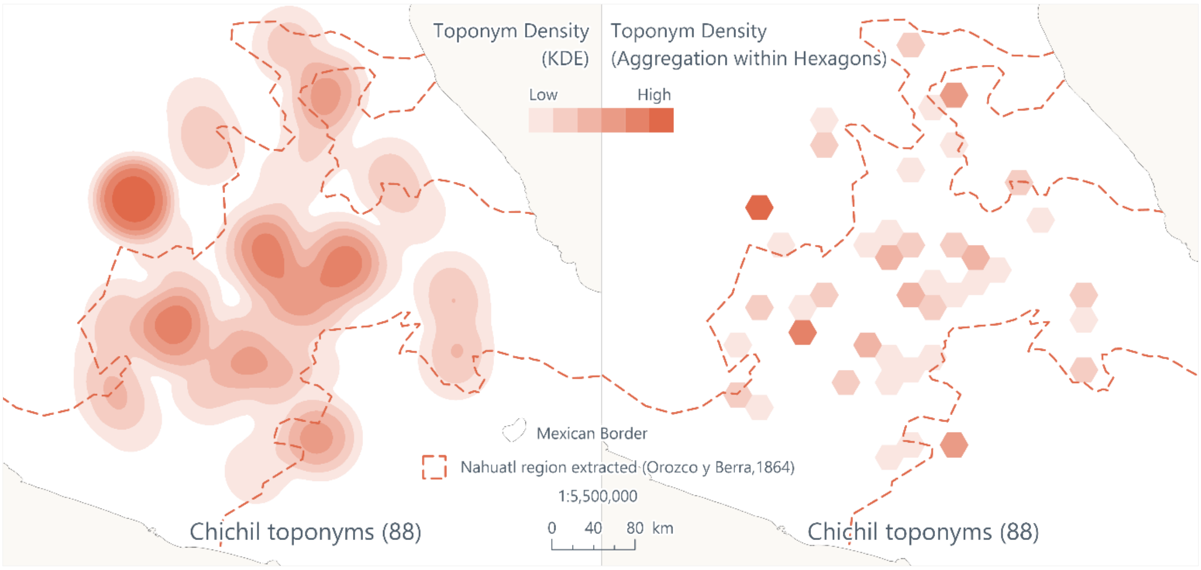

Figure 51 Comparison of the two suggested approaches to analyse toponym distributions

5.2. Other Results

5.2.1. Toponymic Data Exploration

The results from the toponymic data exploration outlined in section 4.1.5 show that the absolute density of toponyms in Central Mexico is high, particularly in Mexico City and the state of Mexico (Figure 52). Also, Eastern Michoacán, Southern Chiapas, and areas close to the coast in Yucatán exhibit a relatively high toponym density. The density of toponyms in the Northern part of Mexico is in general lower. Particularly, Baja California, the North-West of Sonora, and the South of the Yucatán Peninsula exhibit a relatively low toponym density.

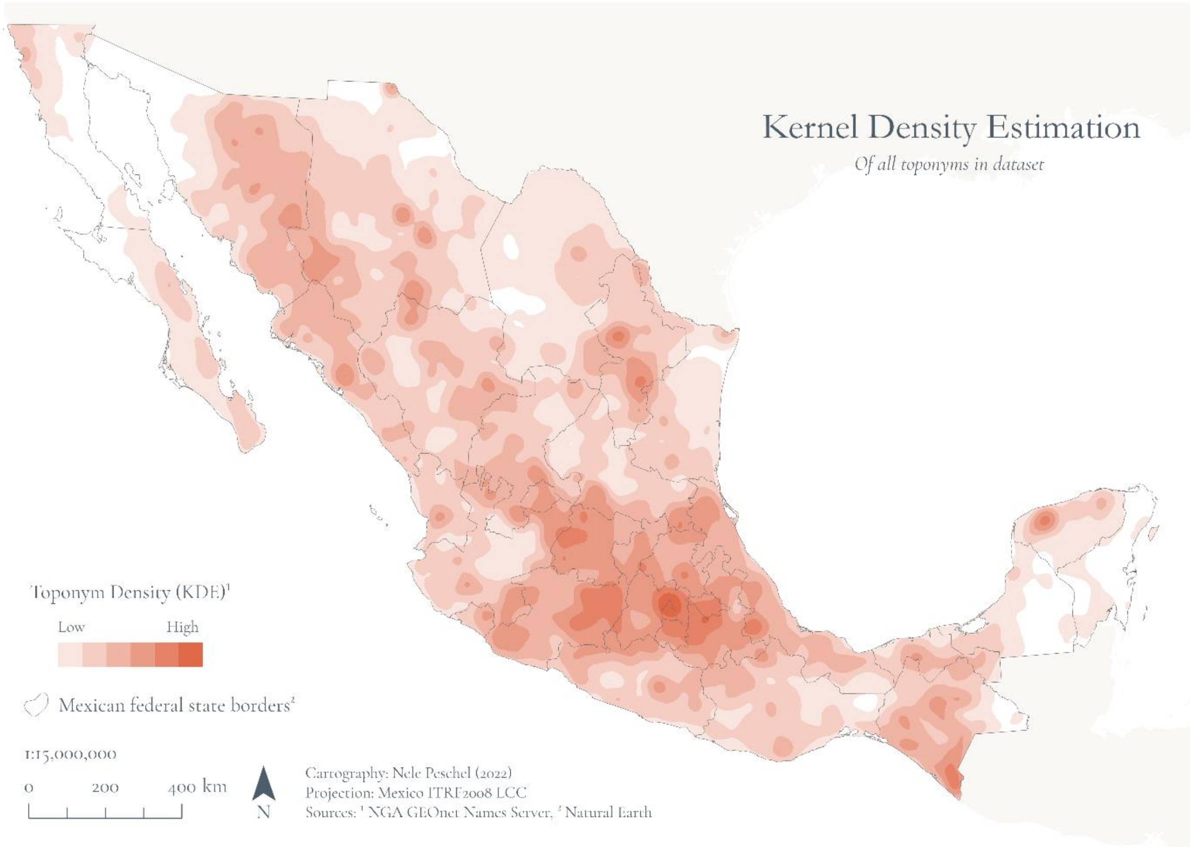

Figure 52 KDE result visualised as isopleths showing toponym concentrations within the study area

The choropleth map visualises the toponym density in relation to the population within Mexican municipalities (Figure 53). The visualisation shows that the concentration patterns are different compared to Figure 52. In general, there is a higher toponym to population ratio in Northern Mexico. Particularly, the East of Sonora, the North-East of Chihuahua, and the North-West of Coahuila exhibit higher toponym density ratios.

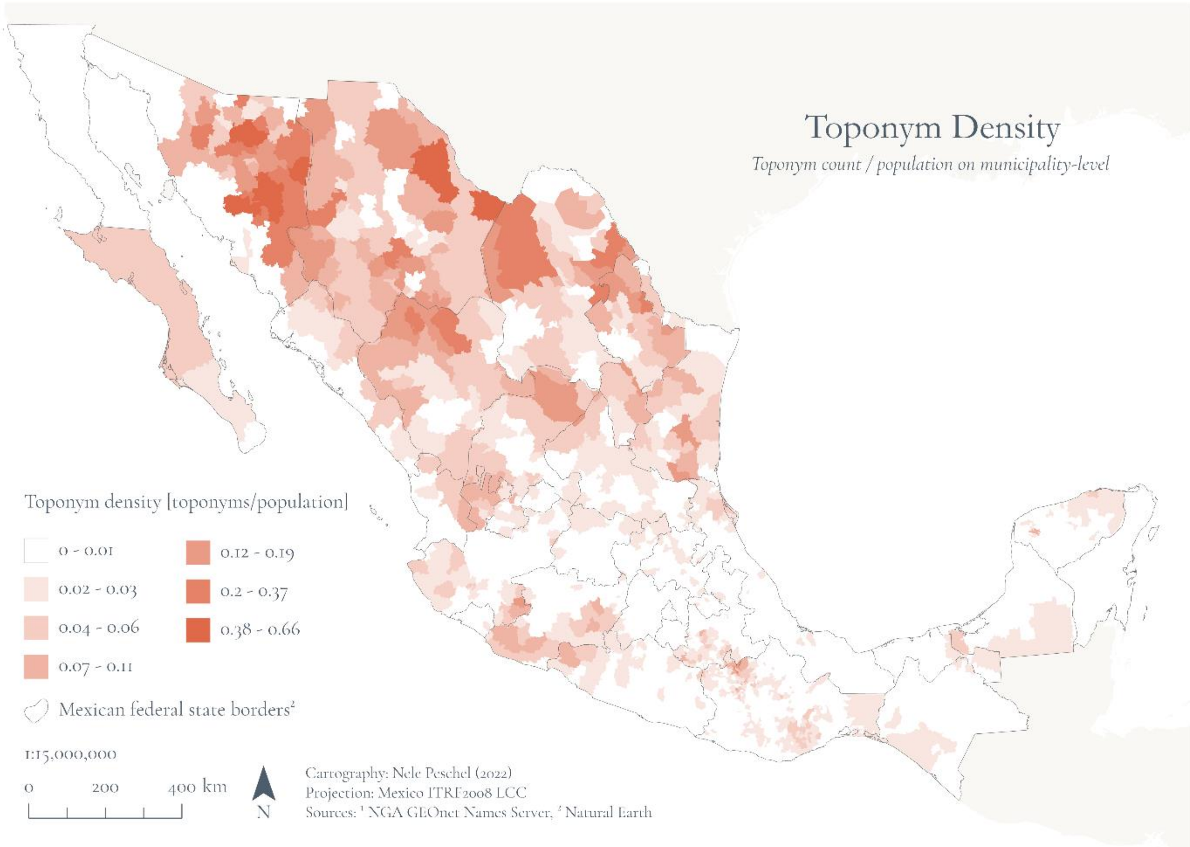

Figure 53 Choropleth map showing relative toponym density within the study area

The results from the toponymic data exploration are useful to get a first overview of the general distribution of toponyms in the study area. It shows that the total distribution patterns of toponym densities are different if we set them in relation to the population.

5.2.2. AoI Identification

Point cluster analysis has proven to be an effective tool to identify AoIs. Two parameters were selected to identify the clusters in a toponymic distribution: search distance and minimum features per cluster. These were selected in an iterative process. Clusters were defined depending on these parameters. The major identified cluster for each of four different toponym groups is shown in Figure 54. The search distances to define these major clusters and the toponym counts within these clusters are presented in Table 13. As these parameters were selected subjectively, the cluster definition was biased. However, this iterative process has shown that the ratio between the selected search distance and the number of toponyms within an identified cluster gives insights into the distribution pattern of a toponym cluster. The lower the ratio, the denser the cluster; the ratio for densely distributed *yucu* toponyms (blue dots in Figure 54) in its major cluster is 0.15 (Table 13). Also, toponyms carrying the morpheme *cuaro* (orange) are densely clustered (ratio = 0.37). The higher the ratio, the more dispersed the cluster; the ratio for more dispersed distributed *chichil* toponyms (green) in its major cluster is 1.36 (Table 13). Also, toponym distributions carrying the morpheme *zoqui* (purple) are more dispersed (ratio = 0.99). The major clusters of each morpheme group are presented in Figure 54. As a rule of thumb, this research suggests a ratio of 0.6 as the threshold to differentiate between dense toponym distributions (< 0.6) and dispersed toponym distributions (> 0.6).

Morpheme group	a: Search distance [km]	b: Toponym count within major cluster	Ratio (a/b)	Dot color in Figure 54
yucu	50	332	0.15	blue
cuaro	100	269	0.37	orange
chichil	120	88	1.36	green
zoqui	150	151	0.99	purple

Table 13 Variables to identify major clusters of different morpheme groups

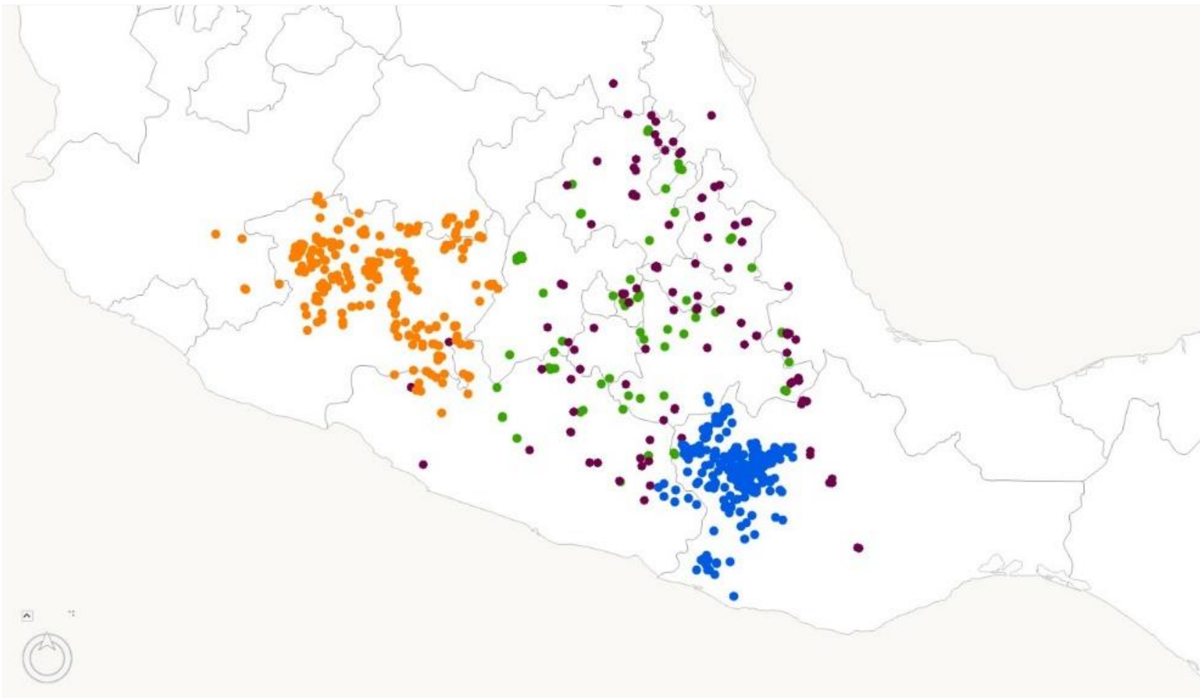

Figure 54 Identified major clusters of different morpheme groups: yucu (blue), cuaro (orange), chichil (green), and zoqui (purple)

For the toponym groups *yucu* and *zoqui* more than one cluster was identified including more than ten toponyms (see Figure 55 & Figure 58). However, large-scale map sheets have not yet been created and constitute future work. Also, cluster analysis was not applied to toponyms of the morpheme group *tepetl* due to their wide-ranging distribution in space (Figure 59).

5.2.3. Overview Maps

The overview maps (Figures Figure 55-Figure 59) are the results of visualising all the toponyms of a morpheme group in GIS. The toponyms were determined and assigned to a morpheme group during the toponym classification. The overview map visualisations show the distributions of selected morpheme groups within the whole study area. The maps depict toponymic patterns on small-scale and function as maps that give an overview of toponym distributions and their spatial patterns. They create the base for further large-scale spatial analysis of toponym agglomerations within a morpheme group (AoI-specific analysis). Furthermore, the maps provide information about linguistic origin, the generic meaning of their morpheme, and the states and regions where the toponyms are mainly located (in the form of text). Overview maps were generated for the morpheme groups *yucu* and *yuta* (Mixtec language), *chichil*, *zoqui*, and *tepec* (Nahuatl language) (Figures Figure 55-Figure 59). It is noted that these maps were largely inspired by the work of Mexican cartographer Sebastián Estremo (Estremo, 2022; Estremo et al., n.d.) (see Figure 6). Furthermore, Figure 55-Figure 58 depict toponym clusters with a red circle, which encloses the toponyms that constitute an AoI (used for large-scale analysis). This visualisation method was also used in the research by F. Wang et al. (2014) (see Figure 13-Figure 14).

Figure 55 Overview map of toponyms of the morpheme group yucu

Figure 56 Overview map of toponyms of the morpheme group yuta

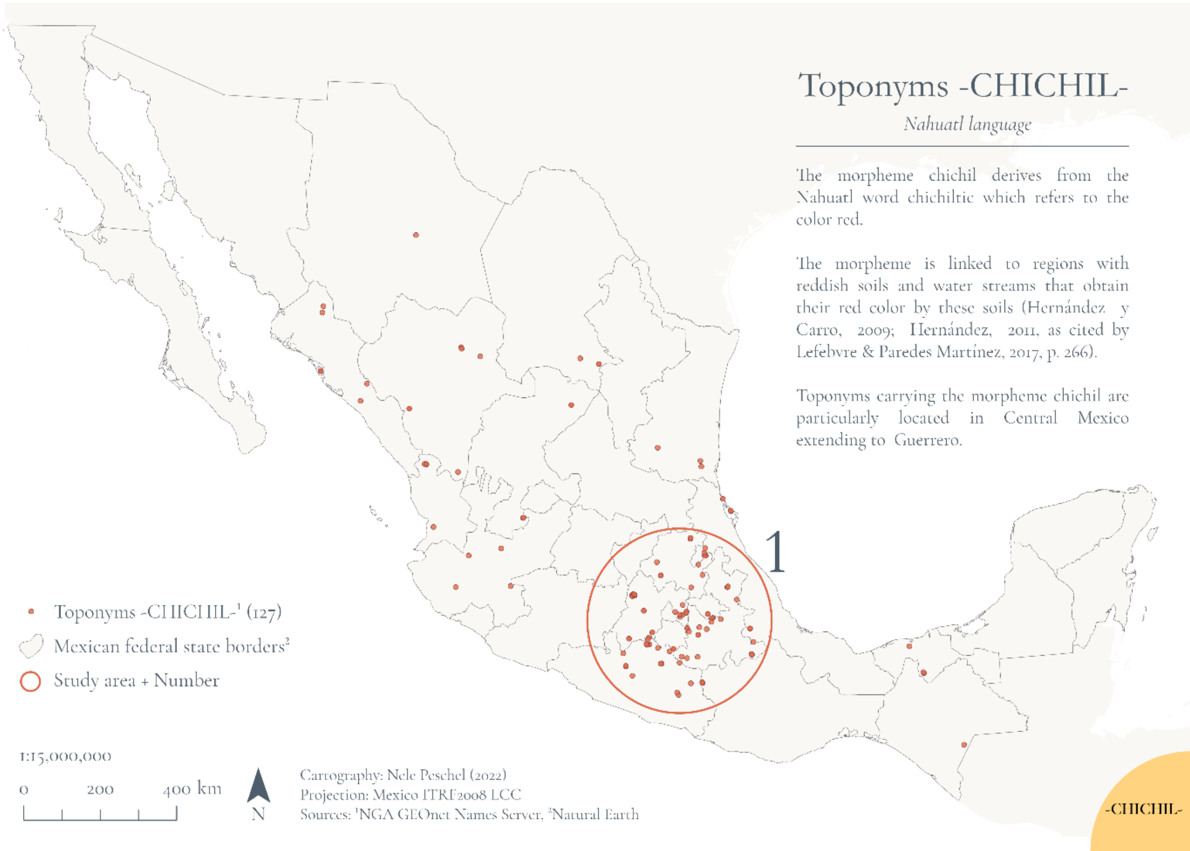

Figure 57 Overview map of toponyms of the morpheme group chichil

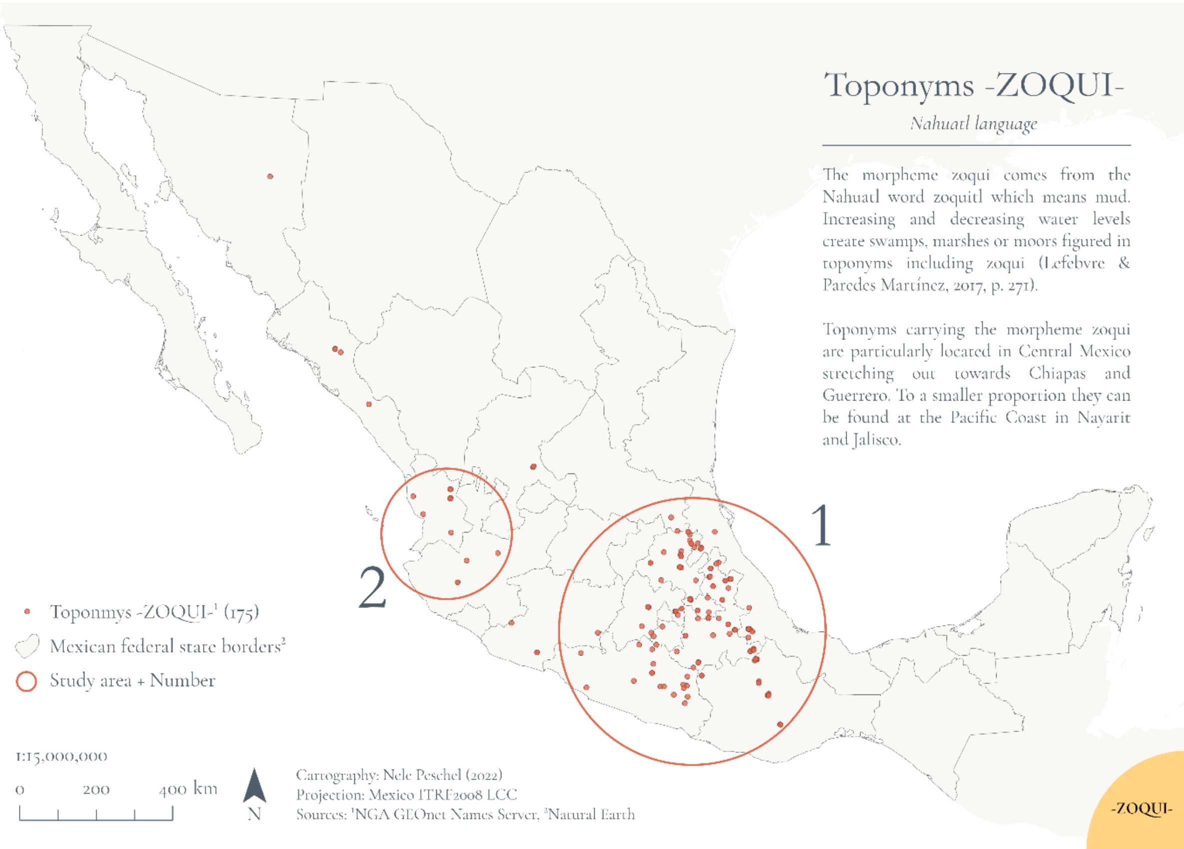

Figure 58 Overview map of toponyms of the morpheme group zoqui

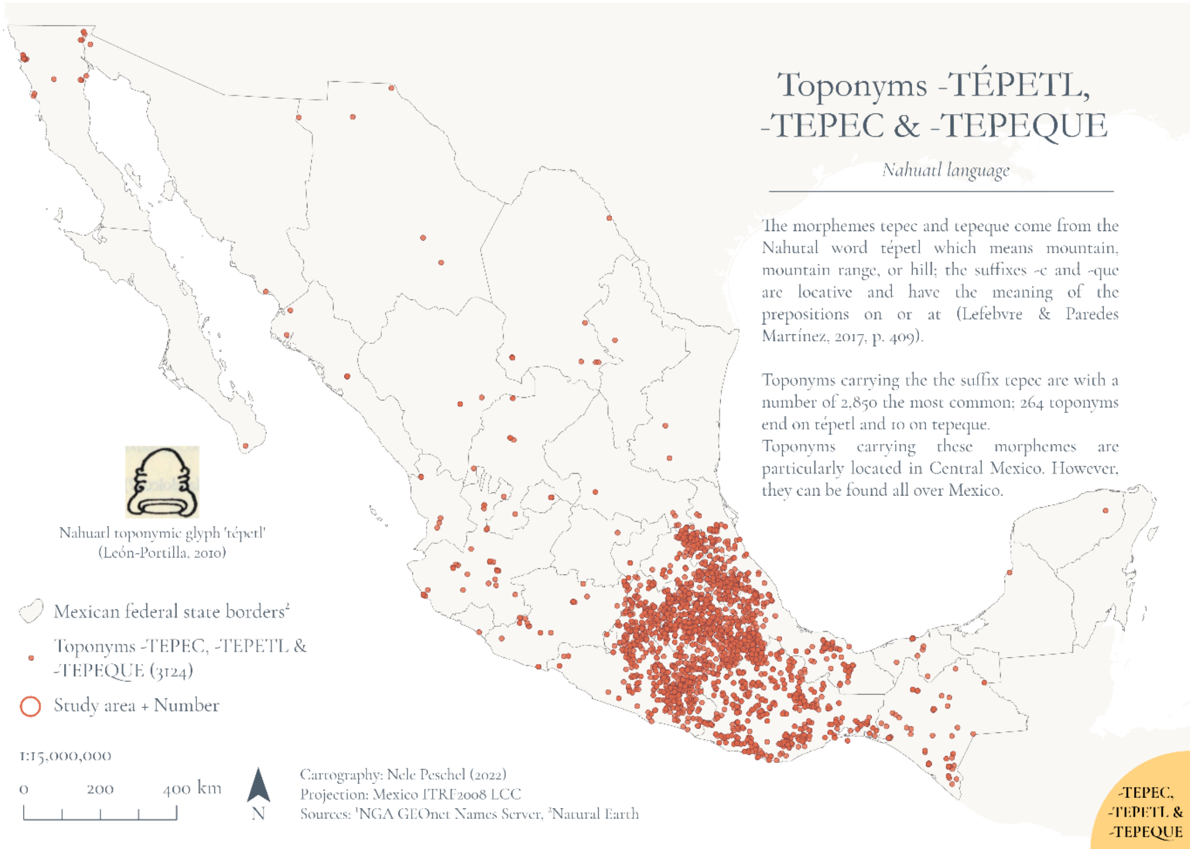

Figure 59 Overview map of toponyms of the morpheme group tepetl

5.2.4. Map Sheets

To assemble and present the results of the methodology applied to Indigenous-derived toponyms in Mexico in a joint layout, this research suggests the creation of map sheets. Two exemplary map sheets for toponyms of the Mixtec-derived morpheme groups *yucu* and *yuta* (both represent their respective AoI number 1) were created according to section 4.3.7 and are presented in Figure 60Figure 61.

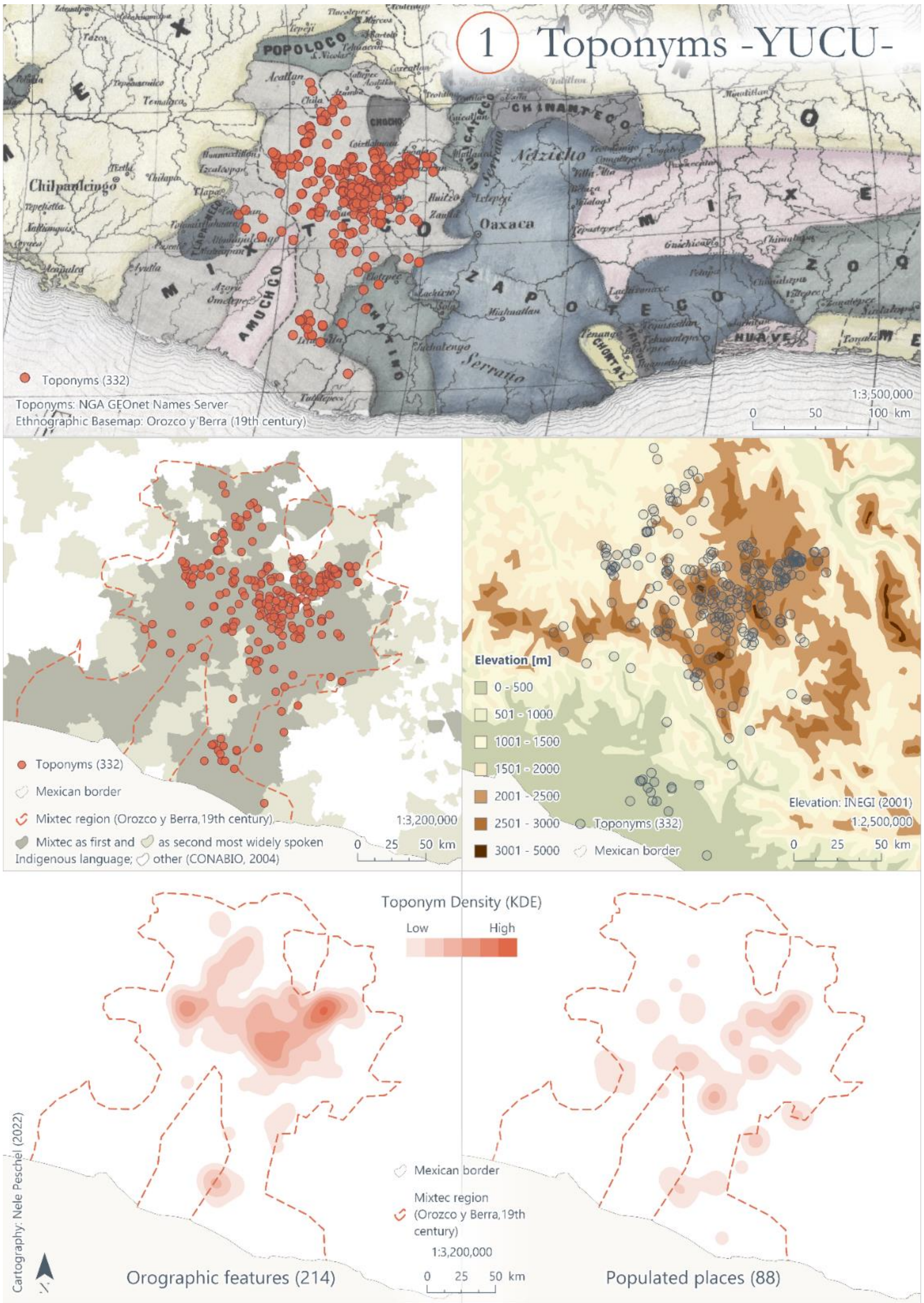

Figure 60 Map sheet of yucu toponyms in AoI number 1

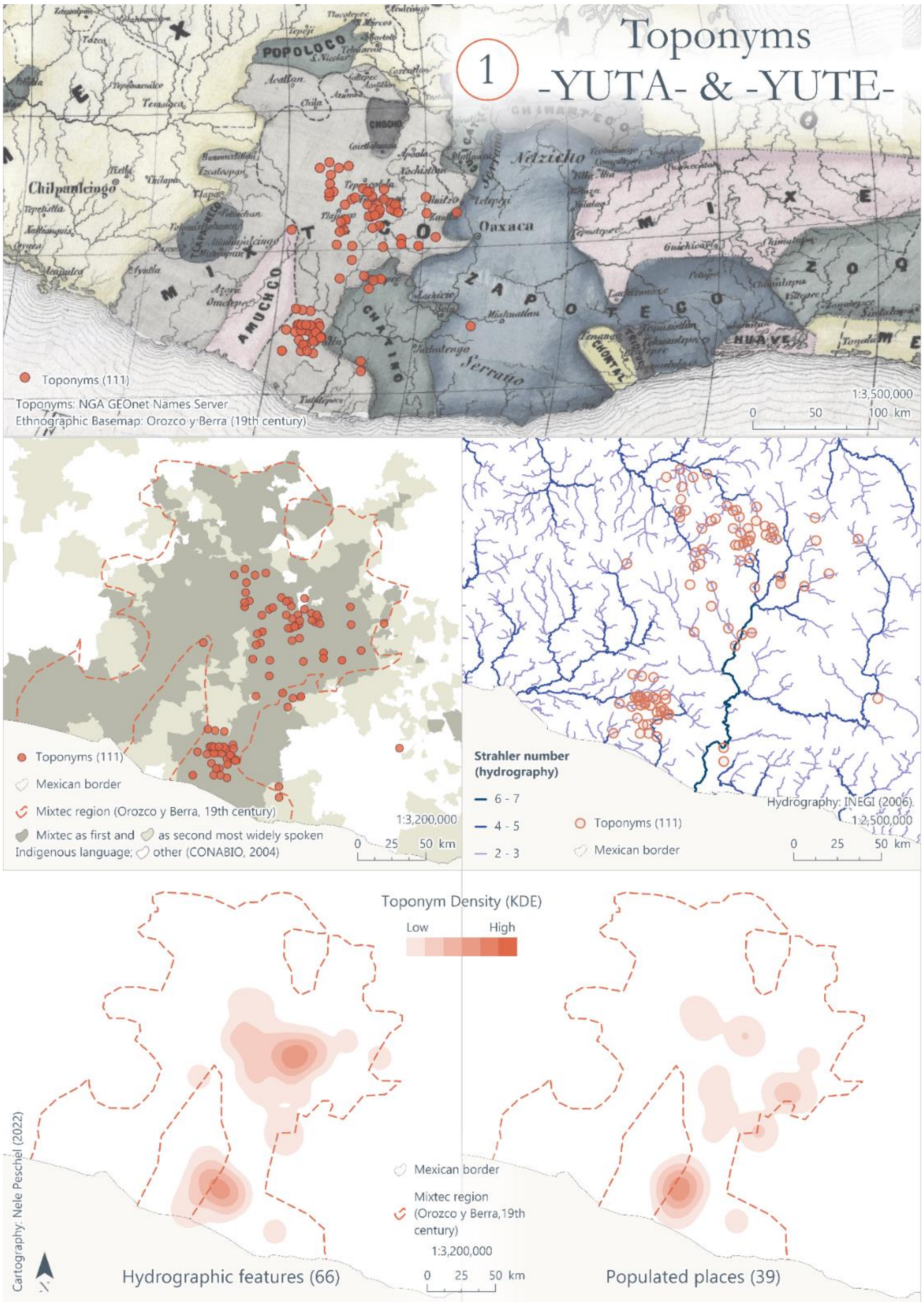

Figure 61 Map sheet of yuta (and yute) toponyms in AoI number 1

5.3. Relevance, Reliability, and Limitations

5.3.1. Relevance

Linguist Cunha confirms that researchers, including linguists, who study the diachronic processes of a specific area can benefit from the application of the suggested methodology to their study area (Cunha, 2022, see section 7.1.1).

Salvador Martínez states that this methodology gives a *"basic vision that is limited in the sense that it needs more analysis, more depth in historical and anthropological terms. However, it is an approach that can indicate how to deepen future analysis. For now, I believe that this methodology is a good advance"* (Paredes Martínez, 2022, see section 7.1.2). Furthermore, he states that the results of this project *"allow us to see the complexity of the subject of toponymy and therefore it is useful for further research"* (Paredes Martínez, 2022, see section 7.1.2).

5.3.2. Reliability

Although morphemes have been selected based on their *uniqueness* and revising third-party literature, it is possible that some toponyms do not derive from the language the morphemes were assigned to.

Furthermore, the generic meaning of toponyms might not any longer refer to the geographic environment they are located at, but will be names given to populated areas or administrative units, such as municipalities, as it is the case for transplanted toponyms (Cunha, 2022, see section 7.1.1). Introducing these transplanted toponyms into the analysis influences the results.

Also, the analysis of spatial relationships is based on parameters defined by the researcher. The results are therefore biased. However, the suggested methodology is useful to indicate spatial relationships. Further research can build on this research's methodology and its findings.

However, the methodology is suggested for a small-scale study area and has proven to be appropriate for extensive research. Therefore, this methodology may be applied preceding to intensive toponymy research. While this methodology was suggested from a cartographic perspective, it may be reproduced by cartographers or GIS experts. However, preceding intensive research may be conducted by experts from other fields including linguists and historian with knowledge of study-area-specific processes.

5.3.3. Limitations

One of the largest limitations is that toponyms of Indigenous origin used by the Indigenous population might not to be found in large toponymic geodatabases (Smith et al., 2012). However, Indigenous place names extracted from, for example, participatory mapping of Indigenous place name knowledge (Smith et al., 2012) could be imported into the database in future research.

Furthermore, each study area exhibits a different toponymy influenced not only by different languages, but also by different historical processes, including different naming or migration processes. In the analysis of the results, these aspects should be considered. This research applies GIS-based methodologies to toponymic distributions and additional datasets with the aim to analyse and visualise their distributions and spatial relationships with language areas and environmental variables. To analyse these visualised spatial patterns in detail, however, it requires interdisciplinary knowledge: *"Other factors that intervene in the toponymic change, in the historical aspects, in the migrations themselves, the loss of the language […] must be taken into account for a more detailed analysis"* (Paredes Martínez, 2022, see section 7.1.2). However, the detailed explanation of spatial patterns is not the aim of this research, but is task for future interdisciplinary work.

Another limitation is concerned with querying toponyms that consist of more than two words. If the data is queried for a suffix or a prefix, toponyms that include the affix in one of their words, which is neither the first nor the last word, will not be detected.

6. Conclusion

This research developed a toponymic classification system to classify a large toponymic dataset by linguistic origin and feature type. The classification was based on study-area-specific literature analysis and feature type information in the original toponymic dataset. Morphemes were used to identify the linguistic origin. They were selected according to the following requirements: deriving from an Indigenous language, influential in the study area's toponymy, and having a generic meaning referring to the geographic environment. Furthermore, feature types were aggregated to new feature type groups according to the generic meaning of the morphemes. The classification scheme based on morphemes and their generic meaning is appropriate for extensive toponymy research. The methodology has proven useful for large toponymic datasets without etymological information. It enriches the toponymic data with information about its linguistic origin and generic meaning by querying the data for selected and meaningful morphemes. The classified toponymic data was used to create overview maps of Indigenous-derived toponym groups in the whole study area using GIS. Furthermore, cluster analysis was applied to the toponym groups to identify large-scale areas of interest and to define whether toponyms were distributed densely or dispersed. A historical ethnographic map was georeferenced and its historical language borders were digitized. The borders were included in the analysis of spatial relationships of toponyms with other variables. These relationships were quantitatively evaluated based on intersections with and distances to the corresponding features – language distributions and environmental variables. Furthermore, two approaches were used to analyse distributions of toponyms of different feature type groups and to compare them in relation to the overall toponym subset. Both approaches calculated (or estimated) toponym densities. The first approach, including KDE and isopleth mapping, was applied to dense toponymic distributions. The second approach, including toponym aggregation within hexagons, was applied to dispersed toponymic distributions. The results were joint in map sheets with a common layout. These map sheets offer the visual baseline to analyse toponym distributions and their spatial relationships with language distributions and the geographic environment. This research has shown that selected toponyms in the study area are mainly located in areas where the language they derive from was and is spoken. Furthermore, it has shown the spatial relationship of a toponym's generic meaning (based on morpheme) with its geographic environment. To review and assess the results of this research, experts from the fields of linguistics and history were interviewed. This research opens possible directions for future research including intensive toponymy research to explain identified spatial patterns and relationships in the study area, the application of the suggested methodology to other study areas, and the extension of the methodology by adding other aspects to the analysis to improve the analysis of toponym relationships.

References

Abdikhalikovna, K. F. (2020). Principles of Toponyms (Place Names) Classifications. *International Journal of Multicultural and Multireligious Understanding*, *7*(6), 73. https://doi.org/10.18415/ijmmu.v7i6.1738

Biblioteca Digital Mexicana A.C. (n.d.). *Manuel Orozco y Berra (1816-1881) Carta etnográfica de México*. Retrieved June 1, 2022, from http://bdmx.mx/documento/carta-etnografica-mexico

Blair, D., & Tent, J. (2021). A Revised Typology of Place-Naming. *Names*, *69*(4), 31–47. https://doi.org/10.5195/names.2021.2260

Chen, Y.-C. (2017). *A Tutorial on Kernel Density Estimation and Recent Advances*. http://arxiv.org/abs/1704.03924

Chloupek, B. R. (2018). A GIS approach to cultural and historical toponymic research in Nebraska. *Journal of Cultural Geography*, *35*(1), 23–43. https://doi.org/10.1080/08873631.2017.1317182

Clark, B., & MacFadyen, J. (2020, October 7). *Lesson 4: Georeferencing*. Geospatial Historian. https://geospatialhistorian.wordpress.com/arcgis-pro-lesson-2-georeferencing-maps/

Cole, D. G., & Hart, E. R. (2021). The importance of indigenous cartography and toponymy to historical land tenure and contributions to euro/american/canadian cartography. *ISPRS International Journal of Geo-Information*, *10*(6). https://doi.org/10.3390/ijgi10060397

Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). (2004). *"Lenguas indígenas a nivel municipal, 2002". Datos extraidos de Serrano C., E., Embriz O., A. y Fernández H., P. (coord.). "Indicadores socioeconómicos de los pueblos indígenas de México, 2002". INI, PNUD y CONAPO. Pimera edición. México, D.F. Escala 1:250000*. 09/09/2022http://www.conabio.gob.mx/informacion/gis/?vns=gis_root/pobla/asgral/loc2010gw

Cooper, J. C. (2020). *Toponymy on the periphery : placenames of the Eastern Desert, Red Sea, and south Sinai in Egyptian documents from the early dynastic until the end of the New Kingdom* (Vol. 39).

Cui, N.-F. (Ed.). (1999). Mega-dictionary of toponyms of People's Republic of China (Zhonghua Renmin Gongheguo Diming Dacidian). In *Commerical Press* (Vol. 2, pp. 3769–3985).

Dent, B. D., Torguson, Jeffrey., & Hodler, T. W. (2009). *Cartography : Thematic Map Design* (B. D. Dent, J. S. Torguson, & T. W. Hodler, Eds.; 6th ed.).

Editors of Encyclopaedia Britannica. (2017, April 18). *toponymy*. Encyclopedia Britannica. https://www.britannica.com/science/toponymy

Edwards, R., & Holland, J. (2013). What is Qualitative Interviewing? In *What is Qualitative Interviewing?* (p. 127). Bloomsbury Academic. https://doi.org/10.5040/9781472545244.ch-003

el Heraldo de México. (2019, November 16). *Este es el significado de los nombres de los 32 Estados de la República Mexicana*. El Heraldo de México. https://heraldodemexico.com.mx/tendencias/2019/11/16/este-es-el-significado-de-los-nombres-de-los-32-estados-de-la-republica-mexicana-132942.html

Esri. (n.d.). *What is GIS? Geographic Information System Mapping Technology*. Retrieved June 20, 2022, from https://www.esri.com/en-us/what-is-gis/overview

Esri. (2022, January 26). *Thematic Mapping with Hexagons*. https://www.esri.com/about/newsroom/insider/thematic-mapping-with-hexagons/

Estremo, S. (2022). *Tóponimos: Memoria de los pueblos en el territorio* (1st ed.).

Estremo, S., Mora, S., Méleard, C., & Zepeda, S. (n.d.). *Los topónimos son la memoria de los pueblos en el territorio*. Tallersiranda. Retrieved April 14, 2022, from https://www.tallersiranda.com/toponimos

Freitag, U. (1993). Five Selected Main Theoretical Issues Facing Cartography: I. Map Functions. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *30*(4), 1–6. https://doi.org/10.3138/dq3r-34t9-7227-51tx

Fuchs, S. (2015). *Toponymic GIS - Role and potential of place names in the context of geographic information systems and GIS*.

Gartner, G. (2021, March 2). *Is cartography a science?* Esri. https://www.esri.com/about/newsroom/arcnews/is-cartography-a-science/

GEONet Names Server (GNS). (2022). *MX.zip*. https://geonames.nga.mil/gns/html/namefiles.html

George, T. (2022, June 24). *Semi-Structured Interview: Definition, Guide & Examples*. https://www.scribbr.com/methodology/semi-structured-interview/

Geospatial Historian. (2013, September 11). *Lesson 4: Georeferencing*. https://geospatialhistorian.wordpress.com/lessons/lesson-4/

GISGeography. (2022, June 20). *What is Geographic Information Systems (GIS)?* . https://gisgeography.com/what-gis-geographic-information-systems/

Hughes, C. F. (2018). *Perceiving Arctic Landscapes Through the Sentiment Analysis of Inuit Place Names*.

Instituto Nacional de Estadística Geografía e Informática (INEGI). (1990). *"Hipsometría". Extraido de Hipsometría y Batimetría, I.1.1. Atlas Nacional de México. Vol. I. Escala 1:4000000. Instituto de Geografía, UNAM. México.* http://www.conabio.gob.mx/informacion/gis/?vns=gis_root/pobla/asgral/loc2010gw

Instituto Nacional de Estadística Geografía e Informática (INEGI). (2006). *RED HIDROGRAFICA DIGITAL DE MEXICO ESCALA 1:250 000 EDICIÓN 1.0.* https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463598428

Israel, J., Martínez, B., & Pineda, C. S. (2016). Cedros: Origin and permanence of the placenames in a Mexican Pacific island Introduction: toponyms in Mexico. In *Island Studies Journal* (Vol. 11, Issue 2).

Király, G., Walz, U., & Podobnikar, T. (2008). *Georeferencing of historical maps-methods and experiences*. https://www.researchgate.net/publication/256307982

Lefebvre, K., & Paredes Martínez, C. (2017). *LA MEMORIA DE LOS NOMBRES: LA TOPONIMIA EN LA CONFORMACIÓN HISTÓRICA DEL TERRITORIO. DE MESOAMÉRICA A MÉXICO.* www.ciga.unam.mx/publicaciones/

Leibniz-Institut für Länderkunde. (2021). *Kleiner Atlas der Siedlungsnamen Deutschlands*. https://siedlungsnamenatlas.leibniz-ifl-projekte.de/#/

Leighton, J. (1895). *The Unification of London: The Need and the Remedy*. http://access.bl.uk/item/pdf/lsidyv3c0c937d

León-Portilla, Miguel. (2010). *Obras de Miguel León-Portilla. VI. Lingüística*. Universidad Nacional Autónoma de México. https://historicas.unam.mx/publicaciones/publicadigital/libros/obras_leon_portilla/545.html

Lieber, R. (2009). *Introducing Morphology*. Cambridge University Press.

Luo, W., Hartmann, J. F., & Wang, F. (2010). Terrain characteristics and Tai toponyms: A GIS analysis of Muang, Chiang and Viang. *GeoJournal*, *75*(1), 93–104. https://doi.org/10.1007/s10708-009-9291-8

Luo, W., Hartmann, J., Li, J., & Sysamouth, V. (2000). Gis mapping and analysis of tai linguistic and settlement patterns in southern china. *Geographic Information Sciences*, *6*(2), 129–136. https://doi.org/10.1080/10824000009480541

Martina, H. (2017). *Mapping and GIS analysis of place names along the Sonora coast in Mexico*. Department of Geography, University of Zurich.

McKenzie, H. (2022, May 9). *Hexagons for Location Intelligence: Why, when & how?* CARTO. https://carto.com/blog/hexagons-for-location-intelligence/

Mongon, S. (2018, November 27). *4 solutions to common problems when making location data maps*. https://carto.com/blog/solve-common-mapping-problems-with-carto/

Nna, O. U. (2015). Decolonizing place-names: Strategic imperative for preserving indigenous cartography in post-colonial Africa. *African Journal of History and Culture*, *7*(9), 180–192. https://doi.org/10.5897/ajhc2015.0279

Openshaw, S. (1983). The modifiable areal unit problem. In *Concepts and techniques in modern geography* (Vol. 38).

Orozco y Berra, M. (2003). Geografía de las lenguas y carta etnográfica de México : precedidas de un ensayo de clasificación de las mismas lenguas y de apuntes para las inmigraciones de las tribus. In *Alicante : Biblioteca Virtual Miguel de Cervantes*. https://www.cervantesvirtual.com/nd/ark:/59851/bmc9g5z0

O'Sullivan, D., & Unwin, D. (David J. (2010). *Geographic information analysis*. John Wiley & Sons, Inc.

Palmer, M., & Korson, C. (2020). Decolonizing World Heritage Maps Using Indigenous Toponyms, Stories, and Interpretive Attributes. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *55*(3), 183–192.

Paredes Martínez, C. S., & Terán, M. (2003). *Autoridad y gobierno indígena en Michoacán: ensayos a través de su historia* (Vol. 1). El Colegio de Michoacán / CIESAS / INAH, Dirección de Estudios Históricos / UMSNH, Instituto de Investigaciones Históricas.

Qian, S., Kang, M., & Wang, M. (2016). An analysis of spatial patterns of toponyms in Guangdong, China. *Journal of Cultural Geography*, *33*(2), 161–180. https://doi.org/10.1080/08873631.2016.1138795

Qian, S., Kang, M., & Weng, M. (2016). Toponym mapping: a case for distribution of ethnic groups and landscape features in Guangdong, China. *Journal of Maps*, *12*, 546–550. https://doi.org/10.1080/17445647.2016.1201017

Schmal, J. P. (2004). *MICHOACÁN: A STRUGGLE FOR IDENTITY*. http://www.houstonculture.org/mexico/michoacan.html

Silverman, B. W. (1986). DENSITY ESTIMATION FOR STATISTICS AND DATA ANALYSIS. In *Monographs on Statistics and Applied Probability* (pp. 1–22). Chapman and Hall.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2013). *Thematic Cartography and Geovisualization* (3rd ed.). Pearson Education Limited.

Smith, D. A., Herlihy, P. H., Viera, A. R., Kelly, J. H., Hilburn, A. M., Robledo, M. A., & Dobson, J. E. (2012). Using Participatory Research Mapping and GIS to Explore Local Geographic Knowledge of Indigenous Landscapes in Mexico. *Focus on Geography*, *55*(4), 119–124.

Tent, J. (2015). Approaches to research in toponymy. *Names*, *63*(2), 65–74. https://doi.org/10.1179/0027773814Z.000000000103

Tent, J. (2017). Indigenous toponyms in the antipodes: A gazetteer-based study. *Names*, *65*(4), 204–214. https://doi.org/10.1080/00277738.2017.1369743

Tichelaar, T. (2002). *Toponymy and language*.

Wang, F., Wang, G., Hartmann, J., & Luo, W. (2012). 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. https://doi.org/10.1111/j.1475-5661.2011.00471.x

Wang, F., Zhang, L., Zhang, G., & Zhang, H. (2014). Mapping and spatial analysis of multiethnic toponyms in Yunnan, China. *Cartography and Geographic Information Science*, *41*(1), 86–99. https://doi.org/10.1080/15230406.2013.831529

Wang, Y., Ge, D., Zhang, T., & Wang, Y. (2019). The sustainable development of choronymic cultural landscapes in China based on geo-informatic Tupu. *Sustainability (Switzerland)*, *11*(16). https://doi.org/10.3390/su11164302

Włoskowicz, W. (2020). Dissemination and Correctness of Geographical Names. Geographers and Cartographers as Toponym Users and Promoters. *Annals of the Austrian Geographical Society*, *161*, 291–324. https://doi.org/10.1553/MOEGG161S291

Yeginbayeva, A., Saparov, K., Aralbekova, M., Atasoy, E., Kizilçaoğlu, A., & Wendt, J. A. (2016). *THE ROLE OF GIS MAPPING METHOD IN TOPONYMY RESEARCH*.

Zahtila, M., & Knura, M. (2022). Visualizing Point Density on Geometry Objects: Application in an Urban Area Using Social Media VGI. *KN - Journal of Cartography and Geographic Information*. https://doi.org/10.1007/s42489-022-00113-7

Zeini, N. T., Abdel-Hamid, A. M., Soliman, A. S., & Okasha, A. E. (2018). An exploratory study of place-names in Sinai Peninsula, Egypt: a spatial approach. *Annals of GIS*, *24*(3), 177–194. https://doi.org/10.1080/19475683.2018.1501759

Zhu, Z., Zhang, H., Zhao, J., Guo, X., Zhang, Z., Ding, Y., & Xiong, T. (2018). Using toponyms to analyze the endangered Manchu language in Northeast China. *Sustainability (Switzerland)*, *10*(2). https://doi.org/10.3390/su10020563

7. Appendix

7.1. Transcriptions of Expert Interviews

7.1.1. Interview with Evandro L.T.P. Cunha

The interview with Evandro L.T.P. Cunha was conducted in English via Google Meet on August 2nd, 2022 from 15-16 PM CET. During the interview the map results Figure 55, Figure 56, Figure 60, and Figure 61 were shown. Information regarding data sources were given and questions regarding methodology and map content were answered.

E: Evandro L.T.P. Cunha

N: Nele Peschel

Essential interview questions are marked in **bold black**.

Essential comments by interviewee (some of them cited in this research) are marked in **bold aqua**.

Introducing ourselves

(...)

E: Since my main area of research is computational linguistics, I think that there are many ways of analyzing toponyms using computational methods and extensive toponym research. And unfortunately, I think that most of the research in the field – well I wouldn't say this is unfortunate but it's a different approach – is more micro (...). Of course, it's important but **I think that in the field it lacks a little bit of broad and – let's say – quantitative work**. (...)

Showing analysis and visualisation

(...)

E: **It's an interesting visualisation since we see clearly this overlap between the toponyms and the Mixtec region – let's say – as the most widely spoken language**. (...)

E: You relate the morpheme to its meaning and then you choose a dataset to use as a background. (...) It's interesting to see, in this case with the rivers, we see that most of the toponyms are around the rivers. (...)

Interview

N: Do you think that these maps help analyzing the relationship between the Indigenous languages spoken in the region and the toponyms originating from this language?

E: Yes, I do, because I think (...) **when we make this cross comparation between the toponyms and other features in the environment, we gain a lot of information that otherwise, sometimes it's hidden when we only see the toponyms without this environmental context**. (...) Sometimes we see some research that shows, for example, a map and the toponyms (...), but we don't have other information regarding the context in which these toponyms appear. And I think that mostly when we study Indigenous toponymy – but of course I'm thinking about the Brazilian case, but it applies to Mexico also of course – **most of the toponyms of Indigenous origin relate to environmental information**. So, I think (...) this is relevant because this is a way of **joining these two information, the name information and the environmental information**.

N: Do you think that these maps could support to indicate also a change of language in the region. By analyzing this relationship of toponyms and language do you think this research could help to indicate whether the people [speaking this language] have migrated to another place?

E: Maybe, but this is a difficult situation because **migration doesn't mean always that names will change, and name-changing doesn't mean always that there was a migration**. It might indicate this (...), because for example when we have a map like the one, the second map () (...) we have here Mixtec as first language in the gray area but as second language in the not so gray area (...) we have for example one isolated dot in a white area, so it's (...) not even the second most spoken language there. So, someone could suppose in the past, Mixtec was the first language over there and then these toponyms survived but then language changed. **Language change also doesn't mean always a migration process, because people can change the language that they speak but it doesn't mean that they have migrated from somewhere because of language contact**. If we have a contact between two languages, one of the effects might be one language influencing the language of the others. We have several examples, for example, during colonization (...). Let's think about the case in the Americas, so they are not Europeans that lived in the Americas but they are Indigenous people who shift to an European language and the case like this one, one might think that this in the past this was a Mixtec region or first language Mixtec, but it could be another reason as well. It could be the effect of the contact, for example of commercial contact between here we have regions not far away from Mixtec as first language region and it could be the influence of this but for example here we see very clear this overlap between toponyms in Mixtec and Mixtec as first language. (...) But in any case, okay so you asked something that I'm not sure we can always answer like this saying that this was caused by migrations. But on the other side **it was caused definitely by some**

kind of language contact. Language contact can be the effect of migrations but can be the effect of other kinds of contact as well, of commercial contact, or previous contact a kinship, relationships, familiar relationships. (...) I think that some kind of contact there is, not necessarily migration but some kind of contact, yes. (...) There is another (...) important point in this discussion. This is the case in Mexico. (...) Of course each situation is different. (...) I give the example in Brazil: (...) maybe it's a specific situation that **Indigenous toponyms were adopted, not by Indigenous people in the 19th century, in the beginning of the 20th century, so (...) we have several towns, not only towns, but also neighborhoods, sometimes big cities, named in Tupi but they were not named by Tupians, they were named by Portuguese (...) in the end of 19th century. So, it's romantic period, people find to value the country. (...) There was this practice of naming places in Tupi, naming families and babies with Tupian names, but by Portuguese people.** (...) If we did this overlap in Brazil, (...) it would be very interesting to see, because we would have toponyms far away from Tupian language zones and it doesn't mean that Tupians migrated there or even they had contact there, but it [Tupi language] was elected as the let's say original Brazilian language – but of course it was not the only one – (...) and then places far away from Tupian regions were named by Portuguese cities, families etc. (...) after Tupian language, using Tupian names. (...) But of course, it's a specific situation. What I'm trying to say is that of course the maps give lots of information but **we also need to consider historical information as well since we are dealing with diachronic data**. So, this data is not synchronic. It's not the data from one specific period but these names were given in different periods of time by different people. (...) Maybe this region was named very far away in time, sometimes before and sometimes a few hundred years later. (...) It's something to consider.

N: [To consider] not only the spatial component, but also the temporal one.

E: Yes, definitely.

N: **And you were saying that some names that maybe come from Tupi language were not named by Tupi people. So, would you say that (...) it's more appropriate to name these toponyms then Indigenous-derived toponyms instead of Indigenous toponyms?**

E: Yes, this is how I like to call them. I use the term – in Portuguese it would be ***baseada em línguas indígenas* – so *based on*; *derived from***. Yes, it could be. But this is the Brazilian case. I think that in Mexico it's very different (...) because in Brazil we still have Indigenous languages widely spoken but mostly in the Amazon. (...) For example, in the South of Brazil it's very rare. So, in Mexico it's more widespread the speaking of Indigenous languages until today. So, it's different because I think that for example let's say that if we have a new village nowadays in a region that speaks an Indigenous language it will receive this name in the Indigenous language. In Brazil probably not. In Brazil, new towns are given in Portuguese names. It's very rare (...). It's a different context but probably in Mexico, and here we see clearly this overlap between the region and the toponyms, between the language and the toponym. **Here it's clear that there is some relationship with the language**. (...)

Showing density maps on the bottom of yuta/yute analysis map sheet

N: Do you think that these density maps could also help comparing the quantity and the extent of toponyms of different feature types? (...) We have toponym groups of all the *yucu* morphemes and then we would divide it in more subsets based on feature type and do you think that using these densities makes sense to compare those quantities and the extent as well?

E: Yes, probably yes, but for example in this case we see that the density of toponyms with hydrographic features and populated places are mostly in the same region, (...) in the South here. But what does it mean? Probably it means that this region has a high density of toponyms in general.

N: That's right. So, you would say maybe you would have to compare it as well to the overall kind of feature types in general in this region?

E: Yes, because otherwise it is just saying that this region has a lot of toponyms let's say in the corpus but it's not clear if it's related to hydrographic features or populated places. On the other side, we see a difference in the other regions, we see that other cluster in the hydrographic features that we can't see in the populated places. (...)

N: I can also show you this one again with another morpheme *yucu* which means hill and there you also see some of course similar patterns, but different concentrations and also the quantity is different.

E: Yeah, but here (...) the density is in absolute number or it's relative?

N: (...) It's based on the overall toponym subset of all yucu morphemes. I used the 332 toponyms and did a density analysis on those and then I took a method to classify these and these class ranges I applied them to the subsets in order to make it comparable. So, these have the same class ranges to make it comparable.

E: Yes, it makes sense because here we see the regions are different. I mean we see for example one of the most – let's say – the highest density in the left in the orographic features (...) we also see it as a highest in the populated places but not as high. (...) And, for example, we see another in the populated places (...) like in the center of the map (...) and **we don't see these in the orographic features as red so it's more relevant what I understand, this region is more relevant for the populated places than for the orographic feature toponyms**.

N: And it could be as you said before that there is more populated place in general not only carrying this *yucu* morpheme.

E: Yes, I think this is interesting. It [density maps] is definitely more difficult to get the idea than the other ones [other maps], but **it's interesting to compare the density of toponyms with different characteristics**.

N: So, you think it is useful to compare these differences between different feature types? It is a useful method or a useful visualisation?

E: Yes, I do. **I think one of the interesting topics in toponymic research is to analyse different features of the toponyms related to the context or to the environment where they appear and I think that this is a nice visualisation for this**.

N: Could you give reasons or explanations from your perspective as a linguist for the spatial patterns of this toponym group by looking at the different maps? (…) You were actually already saying that it's also due to the environment, but these patterns, could you give explanations why they are located here or does that in your opinion depend more on the toponymy of this region in general?

E: Yes. Of course, I don't know the region and I don't know the tales of it. But what I understand is that you – could you go to the map below? Just to remember (…) *yucu* means *mountain*, (…) or *hill*. Could you show the map [Figure 43]? (…) Yeah, we see these high concentrations (…). Of course, it is obvious that usually we have toponyms with specific meaning in the situation, (…) in the environmental context where they make sense. So of course, here we have lots of *yucu* [toponyms] close to the mountains, but we also see (…) low elevation regions with this morpheme. Could you go to the map [Figure 45] below? (…) This is, where we see mostly populated places over there. (…) Because I was expecting (…) this cluster in the South to be not related to orographic features, but to populated places.

N: (…) It might make sense to mention again that *yucu* also means *hill*. So, maybe it doesn't only refer to high mountains but rather on the inclination or difference in height so maybe there is a region with some hills, but of course this must be then studied more in detail. (…)

E: Yes, it could be. (…) It doesn't mean always that the toponym will receive the name or will be named after, like a mountain or a river. **We could have something called river but very far away from a river for many reasons**. For example, there is the phenomenon of **transplanted toponyms**, so they are taken from another place sometimes because – this is usually related to migration – because people from a specific place name the new place where they are living after the place where they lived. So, for example, I was thinking that maybe we could imagine if we had more information on this that some people from the highlands could have migrated to (...) these low lands close to the coast (...). Maybe there was this migration situation, and these **people went from highlands to the coast and then they named the place where they are living at the Coast after the place where they lived in the highlands**.

N: Like in Southern Brazil where a lot of European migrants named places in Brazil like places in Europe, right?

E: Yes exactly, this is very common. And it could have been something like this but not necessarily. It could be also something more directly related to the environment, too. **From a linguistic point of view, the most obvious thing to consider is the relationship between the meaning of the toponym and the context where it appears**. But also, it's important to consider other information, too. If we have historical information on migrations and contact. But here we see, for example, we have some clusters of toponyms in the – I will call it the brown map [elevation map] – sometimes very close to the mountains, for example in the North we see some toponyms, there is this (...) curve and the toponyms are in the middle of them, so places close to the mountains maybe see the mountains.

N: That's right. It doesn't have to be on the mountain but maybe in a region where you see them, where you have the perspective on the mountains (...), like these for example, these two places, and maybe these are then populated places rather than orographic features? (...)

E: Yeah, maybe. Let's try to see in the toponym density map. (...)

N: This is actually orographic features, and we were assuming that it could be rather populated places. But it is not like that. But maybe there is some examples where it could be like this [that toponyms carrying the meaning *mountain* or *hill* found in lowland areas are rather of the feature type populated places than orographic features] and this could be a reason to explain but it not necessarily the reason. This is where you said you have to consider also other context information, not only the geographic environment. (...)

N: **Do you think that these maps could be reproduced easily with other toponym groups, for example, also in Brazil that are classified by linguistic origin?**

E: Yes, this is what I was trying to think. **I think its's very beautiful and I think this is something that toponym research should invest more – in visualisations and beautiful maps**. I think that yes definitely it could be used in other context if we have the data, if we have the information, yes. (...) If we think about Brazil, one interesting topic that came to my mind is to try to do the same using Tupian names and relating [them] to of course the environment but considering that we know that there was this process of

Tupian naming not by Indigenous people. So, try to see if at least some connection with the environment is present in the naming of these places.

N: So, also the naming-practices have to considered or studied or the *how* or *who* did name the place.

E: Yeah, what I was thinking for example was to do the same for places in Brazil that carry Tupian names in the South-East region where we know that there was this process that I told you and try to see, for example, in Tupi *ita* means *stone* but it's used for *hill* also – hills made of stone. Are these places (...) using this morpheme *ita* (...), can we relate them to the orographic environment? So, to the mountains, to the hills? Because one question that I have – this would be a research question – was this place just named randomly using (...) Tupian morphemes or where they named really considering the environment where they are? So it's like, for example, if I try to name something in my city using a German name for example, will I just put the name that I like or...

N: ...according to something that it reminds me of? Yeah, that's interesting. Very difficult to study this research question but very relevant.

E: Yes, because we would need some extra information on who named the place. But usually, for example, the towns they were named by Europeans – towns in São Paulo or Minas Gerais – they were named not by Indigenous people. So, they used the names in Tupi but they were not – definitely not – named by Tupians. This is different from Amazon. In the Amazon it is a totally different situation. But in the South-East of the country, it's like this.

N: And maybe you would also see patterns or clusters of places that were actually really named by Tupian people and clusters of Tupi place names that were named by Europeans?

E: Yes, maybe. (...) **Actually, what I feel from your work is that if you'd have the data, (...) I mean the corpus of toponyms and the maps, you could create something like this for any place in the world I think**.

N: And you were starting by saying it's beautiful (...). Would you say that this reproduction could also be practiced by linguists, or would you say that maybe rather cartographers would be the person to create this kind of maps or do you feel that linguists could also be involved in the map creation?

E: Yes, **I think definitely linguists could be involved in the map creation**. And I don't know how you create this or if you created a software for this or if you just put the data and the map and it creates. But if so, yes I think so. I don't know how technologically difficult it is for creating the maps of course, to create the first maps, definitely it's more difficult but I don't know if you have some way of creating a software that we could just put the data, or for example a script...

N: ... you just run it and then you would have this output...

E: Yes, exactly. If you'd have this, I would definitely try to use it. (...)

N: **Which limitations do you see with these maps?** But maybe we actually have already discussed this. You gave a lot of (...) [input] on context, maybe some more context could be also mapped or involved in order to assess it or evaluate it more to make it more meaningful.

E: Yes, in general what I think is that **the map gives lots of information, but it also needs lots of information to be interpreted so I think this is the limitation of any map I think that we need to understand it correctly.**

N: Side information, not only the visual representation but also of course context information in the form of text, for example.

E: Yeah, I think that these resources could be very rich for researchers that study a specific area and know the diachronic processes in the area.

N: **So, do you think that these maps they could directly help linguists to connect thoughts and with their knowledge on naming practices they could benefit from these visualisations in your opinion?**

E: Yes, I do. Definitely yes. (...)

N: **Do you think that now that you have asked me these questions on the methodology and we discussed, it became clearer to you than just seeing the maps itself?**

E: Yes, now I understand more after we have discussed and you explained me I understand more, the idea behind and how you created them.

7.1.2. Interview with Carlos Salvador Paredes Martínez

The interview with Carlos Salvador Paredes Martínez was conducted in Spanish via Google Meet on August 11th, 2022 from 18-19 PM CET. During the interview the map results Figure 55, Figure 56, Figure 60, and Figure 61 were shown. Information regarding data sources were given and questions regarding methodology and map content were answered.

C: Carlos Salvador Paredes Martínez

N: Nele Peschel

Essential interview questions are marked in **bold black**.

Essential comments by interviewee (some of them cited in this research) are marked in **bold aqua**.

Introducing ourselves

C: En mi caso yo soy historiador. Sin embargo, desde un principio entré a trabajar a un centro de investigaciones en antropología social. De esta manera fue que me introducí en la etnohistoria que es más bien la visión histórica a través de los documentos con interrogaciones y cuestionamiento de antropología social de manera que hacemos etnografía del pasado, no con poblaciones actuales, sino con poblaciones del pasado. De esta manera ahí en la etnohistoria tratamos de acercarnos a los pueblos indígenas del pasado a través de las distintas disciplinas que abarcan los pueblos indígenas, desde la historia, la antropología, la lingüística, la geografía también. Varias disciplinas tenemos que abordar para tener una visión lo más completa posible del pasado indígena, de México en este caso. Y así, a través de muchos años que ya he permanecido en este centro de investigaciones, el Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), me ha permitido hacer investigaciones regionales tanto en la entidad de Puebla, como en Michoacán. Entonces he abordado distintos temas de los pueblos indígenas desde el siglo XVI y eso mismo también me ha llevado a tratar de entender un poco la época prehispánica sin ser arqueólogo, pero si trato de estar enterado en lecturas de arqueología para ver ese tránsito entre la época prehispánica y la colonial. No se trata de un corte completamente, que se olvida todo en la época hispánica. No es así, sino hay momentos de transición y cambios que nos explican algunas cosas del pasado. Haciendo la comparación de Puebla y Michoacán me doy cuenta, que desgraciadamente para Michoacán tenemos muy pocas fuentes, en Puebla tenemos una gran cantidad de información. Sin embargo, para Michoacán se han destruido muchos archivos. El trabajo arqueológico, si hay importantes trabajos, pero solo hasta recientemente se están haciendo de manera que en el caso de Michoacán carecemos de esta información. **Los códices prehispánicos por ejemplo si los hay para las**

zonas de Puebla y Oaxaca, pero para Michoacán no tenemos, no hay estos códices prehispánicos. Entonces hay una gran carencia para poder explicar cómo era la situación en la época prehispánica. Este es uno de los principales problemas. Pero en fin ya últimamente justamente conocí a Karin Lefebvre cuando vino a Michoacán. Ella vino con intenciones de hacer una estancia postdoctorado después de su doctorado en Francia y yo la conocí a ella por un congreso de americanistas. Entre sus proyectos de investigación estaba justamente hacer un coloquio sobre toponimia. Ahí fue que ella me invitó a formar parte de la organización y ahí coincidimos en muchas cosas y creo que fue un buen proyecto que cristalizó con esta publicación (see Lefebvre & Paredes Martínez, 2017). (...)

C: Lo interesante de esta publicación es la diversidad de enfoque. Son distintas disciplinas que abordan solamente el tema de la toponimia, ésta es la riqueza de esta publicación. (...)

Showing analysis and visualisation results

(...)

C: Sobre Oaxaca hay una gran cantidad de grupos étnicos, de lenguas distintas. Así, uno de los fenómenos de estudio es la perdida de la lengua y consecuentemente la perdida de los topónimos. Es un fenómeno muy generalizado que están experimentando en todo México y lógicamente en Oaxaca también. (...)

C: Esto es el cambio del topónimo mismo. Sufre muchas modificaciones y si lo vemos a través del tiempo, evidentemente que hay muchos factores que pueden cambiar, el hecho de la perdida de la lengua, o bien la influencia de otras lenguas que están en contacto directo con la lengua, es muy claro estos cambios de manera que hay que tomar en consideración estas posibles variantes del topónimo mismo. (...)

C: Desde luego la fuente que estás utilizando del mapa del Orozco y Berra es fundamental porque es uno de los más antiguos que dan un panorama de las lenguas en México en el siglo XIX pero si hay que ver estos cambios que a partir de entonces evidentemente han sufrido las poblaciones. (...)

C: No solamente desde la conquista de México es que aparecen los nombres cristianos, sino también actualmente el fenómeno de la migración de Estados Unidos y del entorno están introduciendo cambios que son claramente ya muy posteriores del siglo XX y XXI por la cuestión de la migración y las influencias norteamericanas por ejemplo cambiar sencillamente los topónimos por un topónimo como Florida. Ya nos lleva a otra situación. Eso es lo que hay que considerar en un estudio sobre toponimia. (...)

N: ¿Cree que estos mapas ayudan a analizar la relación entre la lengua indígena que se habla en una región y los topónimos que se originan en esta lengua?

C: Si, como te decía **el mapa de Orozco y Berra es una fuente muy buena para tener una perspectiva en un momento dado. Es una base importante para considerar.** Entonces más bien yo te preguntaría ¿cuál es la comparación que haces entre el mapa de Orozco y Berra y las derivaciones hacía después? ¿Cuál es tu fuente para hacer estas comparaciones?

N: (...)

C: Ahí el cambio toponímico es un tema fundamental. Pero me imagino que hay otros temas que estás abordando en tu estudio de toponimia. ¿Qué otros temas o subtemas estás abordando?

N: Un tema importante es obviamente el significado del morfema y analizar la relación entre este significado y el entorno geográfico a través de mapas, a través de visualizaciones espaciales como muestra este mapa, aquí a la derecha (Figure 44) quiero analizar de forma visual esta relación, pero también pienso en cuantificarlo y analizar cuántos topónimos están a una distancia de 5 kilómetros del sistema hidrográfico, por ejemplo. (...) Y ahí le quisiera hacer la pregunta de si **¿Ud. cree que estos mapas también ayudan a analizar la relación espacial entre el significado de los topónimos y el entorno geográfico en los que se encuentran estos topónimos?**

C: Ahí hay un aspecto interesante en cuánto a los topónimos. **Lo más conocido son las poblaciones, las ciudades, los asentamientos humanos. Pero detrás de estos, dónde hay asentamientos humanos que evidentemente es lo que ocupa mayor interés sin embargo hay topónimos para designar un paraje, una parcela de cultivo, un terreno sencillamente, de manera que estos topónimos nos pueden significar mucho en el sentido del uso del agua, en el sentido del uso de los bosques, por ejemplo**. (...) Aparentemente alguien que acude el terreno y que es ajeno **al lugar, no significa nada, pero localmente la gente local sí tiene mucho significado, porque ahí significa que se acumula el agua o que ahí es una zona muy fértil**. (...) **Son territorios que aparentemente no nos dicen nada pero que tienen mucho significado localmente y en la lengua esto se nota claramente**.

N: En el idioma se refleja. Entonces el topónimo indica algo acerca del entorno.

C: Así es. (...)

C: El trabajo de Juan Gallardo Ruíz que es sobre los purépechas actuales en el que la gente le da un significado a la zona de los bosques que son muy acosados por los talamontes, los que cortan los bosques y le dan un nombre específico al predio, al lugar, indicando justamente este temor por la tala de los árboles, o sea son fenómenos que están actuales hoy en día y que hay que considerar en los estudios de toponimia.

N: Es entonces una disciplina muy rica, muy amplia, con muchos puntos de investigación.

C: Así es. **No solamente es el campo lingüístico el que nos da el significado sino otros fenómenos. Por eso es la riqueza de los topónimos.**

N: Le tendría otra pregunta. **¿Si Ud. cree que estos mapas podrían servir para indicar un cambio de uso del idioma en una región?** Ahora hablamos de nuevo de idiomas, de lingüística, pero quizás tiene un pensamiento acerca de esto.

C: Desde luego **está el problema de la pérdida del idioma** porque un tema muy específico del trabajo que te voy a mandar de una colega que es chatina de Oaxaca. Ella está interesada en los temas de la toponimia también, menciona que en reuniones que tiene con la genta ahí **en las poblaciones donde se habla chatino la gente duda de la persistencia de los topónimos en lengua chatina porque la gente ha perdido el uso de lengua**. **Entonces la gente misma dice ¿para qué mantener el nombre de tal o cual pueblo, de tal o cual lugar si ya no hablamos la lengua?, ¿si ya no está vigente la lengua y que ni siquiera sabemos lo que significa? A estos grados llega la perdida de la lengua y las consecuencias de la perdida de los topónimos**. En cambio, la gente mayor tiene una posición todo lo contrario, la gente mayor manifiestan en las entrevistas que no tiene por qué ponerles nombres nuevos, ¿por qué los nombres nuevos de héroes nacionales? O cuestiones así. Dicen, ¿pero si ya tiene nombre este lugar en chatino, por qué vamos a cambiarle el nombre? (...)

C: El mismo Miguel León-Portilla – que tuve la fortuna de ser discípulo de él – hace la crítica de que, **¿por qué le quieren cambiarle el nombre si los nombres reflejan una cosmovisión del pasado, de nuestras lenguas autóctonas? ¿Por qué hay que cambiarles el nombre a los lugares?** Eso es muy evidente, ese problema.

N: Sí, porque es un patrimonio cultural, ¿verdad? Estos topónimos que deberían permanecer en el futuro porque son la identidad de los pueblos de la gente.

C: Así es. (...)

N: Quizás ya hemos hablado de esto, pero **¿Ud. cree que es útil analizar esta relación de la lengua y de la toponimia o es decir la relación entre la toponimia y el entorno geográfico? Y si Ud. cree que eso es útil, ¿por qué?**

C: Sí, **desde luego la lengua es un factor determinante para identificar el origen del topónimo mismo**. Sin embargo, ahí precisamente la primera sección de los trabajos del libro en el que entre el mío, y los de la INEGI, y de Juan Gallardo, **no es solamente atribuirle a la propia lengua, el origen del topónimo, sino la motivación que lleva a denominar a tal o cual lugar de determinada manera**. Y ahí lo que decimos un poco en la introducción son los aportes de un autor español de apellido Trapero, en el que justamente él planteó al igual que el autor mexicano, Ignacio Guzmán Betancourt, diciendo que el origen de los nombres puede variar por el hecho justamente de ser antropónimos en el sentido de tener el origen en tal o cual dirigente, (...) o un Dios, por ejemplo, Pallas Atenea, que es el origen de Atenas, es una diosa (...) entonces el origen es en honor de la Diosa Atenea. Entonces ese tipo de motivaciones. ¿Cuál es el motivo para denominar tal o cual lugar de determinada manera? Evidentemente que, sí partimos de la lengua propia, pero también otras razones (...) que dan origen a denominar a tal o cual lugar de determinada manera. Esa forma de cuestionarse: **¿Por qué o cuando o en qué momento se denomina tal o cual lugar? Hay que abrir el panorama para tener una serie de postulados para identificar distintas razones por las cuales se denomina tal o cual lugar de esta manera**.

N: Entonces Ud. dice que también es esencial investigar el tiempo y la motivación que existía para nombrar o denominar un lugar. Hay que también investigar el ¿Cuándo? y el ¿por qué?, no solo el idioma.

C: Así es. Igualmente, **el manejo del poder que está dominando la región y que está imponiendo un topónimo a pesar de que pueda haber pueblos de distinta lengua**. Esto es lo que yo hago en mi ponencia: **A pesar de que había una población de origen matlazinca en Michoacán, a pesar de esto: los tarascos que eran los que dominaban, imponían sus topónimos de origen purépecha** y también es lo que hace Brígida von Mentz en el artículo que publicó en revista Historia Mexicana que te envié de Guerrero y **migrantes del estado de Guerrero en el que los nahuas, los mexicas, están conquistando toda esta región del Norte de Guerrero, y están imponiendo sus topónimos nahuas – de origen nahua – a pesar de que ahí son chontales y tlapanecos y de distinto origen, pero los nahuas, por ser los conquistadores, están imponiendo sus topónimos**. (...) Están conquistando el lugar y no solamente imponen sus tributos y trabajo, sino también **cambian los topónimos locales, que eran antiguos, los cambian ellos por ser los que están dominando en este momento y esto estamos hablando de la época prehispánica**. (...)

N: Para esto se necesitaría mucha más información en cuanto a ¿quién? y ¿cuándo?, pero como yo estoy haciendo un análisis más (...) a una escala pequeña entonces cree que para empezar a estudiar los topónimos **¿cree que usar esta metodología de buscar la base de datos por morfemas que son quizás no únicas pero muy dominantes en un idioma indígena? ¿cree Ud. que esta metodología es buena para aproximar esta temática? ¿Para tener una idea y para tener también quizás saber dónde se podría después investigar más a detalle en cuales regiones?**

C: ¿O sea tu base de investigación es los topónimos en las lenguas indígenas y compararlas con el espacio geográfico en cartografía contemporánea? ¿Esta es tu investigación? ¿Y esto a escala de toda la República Mexicana?

N: Sí, busco por morfemas en toda la República Mexicana pero como le he mostrado hago estos mapas de sobrevista y luego después del análisis donde se juntan estos topónimos ya escojo un área más a detalle, a una escala más larga para después seguirlo investigando más a detalle. Pero tomo en cuenta toda la República Mexicana en mi metodología.

C: Solamente para que tengas el panorama completo. Haciendo la comparación del siglo IX con Orozco y Berra y la situación contemporánea, el instituto nacional de lenguas indígenas de México ya tiene un censo bastante completo sobre las distintas lenguas habladas hoy en día en México (...). Para tener esa visión de México contemporáneo y sus lenguas que son más de 60 lenguas, pero como con 300 variantes y estas variantes justamente son de Oaxaca y otros lugares de México, (...) La variante es tan importante que entre un hablante y otro hablante no se entienden.

N: ¿Y este censo da información también sobre los topónimos o solo de los idiomas que se hablaban y que se hablan?

C: No, es solamente sobre los idiomas y las variantes. Precisamente, hablando recientemente con Frida Villavicencio que es la colega del CIESAS, que es lingüista y es especialista en purépecha ella es la que coordinó el trabajo que te envié del visor toponímico mexicano y que se encuentra en la página web oficial del CIESAS en el laboratorio de lenguas indígenas mexicanas.

N: Un trabajo excelente, que me ha gustado mucho.

C: Así es. Ella hacía este señalamiento que lo que hicieron en el INALI (...) es muy importante porque es una clasificación de lenguas actuales, pero no llegan a ese grado del análisis toponímico. No es esto. Esta sería la labor que justamente sigue por hacer adelante. (...)

N: Ud. ha hablado de que el tiempo y la persona o el grupo que lo dominaba a los lugares es importante, pero **¿qué piensa Ud. de la relación de la denominación o del significado y el entorno? ¿Cree que estos mapas son útiles para analizar esta relación del entorno junto con el significado?** Ud. ha hablado antes un poco de los topónimos que indicaban si los suelos estaban fértiles por ejemplo quizás platicar un poco más sobre lo que ya ha empezado a discutir.

C: Sí, claro. **Desde luego que sí son pertinentes estas comparaciones y llevar a la cartografía el análisis toponímico, esto creo que es un paso importante lo que estás haciendo**. No sé si has revisado todos los trabajos del libro de *la memoria de los nombres* porque ahí los trabajos de Rocío Hernández es solo un ejemplo, solamente utilizando los códices, algunos inclusive prehispánicos, códices en donde están los topónimos, la iconografía de los códices del agua, el uso del agua justamente, y su topónimo en la lengua propia. Creo que este trabajo es muy importante porque te da idea justamente de la importancia del uso del agua, de las represas, de los conductos del agua, del uso del riego, de la acumulación del agua. O sea, toda esta parte de la tecnología agrícola, ahí está expresada en ese trabajo de Rocío Hernández. Sí, desde luego te abre un panorama. (...)

C: ¿Entonces tú estás tratando de hacer un panorama general de todo el país mexicano?

N: Un poco, como desarrollar una metodología de cómo se podría empezar a investigar los topónimos o la toponimia del país de una forma extensiva teniendo una base de datos amplia para después poder ir más a detalle, sean lingüistas, historiadores o cartógrafos. Esta es la idea que tenía. Y es una temática que me gustó mucho. (...)

N: Quisiera preguntarle si **¿cree que estos mapas de densidad ayudan a comparar la cantidad y también la extensión de los topónimos de diferentes tipos, o de diferentes elementos geográficos, como los topónimos de hidrografía a los topónimos de lugares habitados?**

C: ¿O sea tu identificas topónimos con este carácter de hidrográficos con las raíces de agua?

N: Es buena pregunta. De hecho, no lo identifico yo, sino ya está en la base de datos de qué elemento se trata. O sea, está el nombre y luego cada topónimo esta caracterizado por elementos geográficos, por ejemplo, rio o montaña. Entonces eso ya está en la base de datos y yo lo filtro por estas clases y analicé la densidad y la visualicé de esta forma. (...) **¿Entonces la pregunta seria si estos mapas ayudan a comparar la cantidad y extensión de estos topónimos de diferentes grupos o características?**

C: Yo creo que **si es válido esta comparación y puede ser muy útil para identificar.** ¿O sea tu parte es que estos topónimos contienen esta raíz de montaña, de cerro, y también de los lugares en donde está registrada la lengua mixteca?

N: Lugares dónde hay gente viviendo, más las poblaciones. Sí, donde se habla la lengua es más el mapa de la izquierda que lo muestra. Pero ahí estas poblaciones que se ve aquí la densidad de estos puntos, de estos topónimos de poblaciones, no se sabe si ahí se habla el idioma, sino que son lugares que llevan este morfema que proviene de esta lengua, pero no se sabe si en realidad todavía o si se hablaba en el pasado este idioma.

C: Bueno, pero tu base de información es el mapa de Orozco y Berra, ¿no?

N: Sí, la línea roja es el área en donde históricamente se ha hablado mixteco. Obviamente, son líneas muy exactas, en la realidad no es así. Pero la base de datos es muy amplia, (...) que no se encuentra ahí el origen lingüístico en esta base de datos, si no yo con la búsqueda de morfemas, le indico la raíz lingüística. (...) Porque no existe esta información. (...)

C: ¿Has tomado en cuenta otros topónimos, otras raíces distintas de temas hídricos e hidráulicos? (...)

N: Sí, de hecho, voy a aplicar esta metodología a más, pero obviamente voy a tomar una selección de morfemas porque no puedo investigar todos los morfemas en su totalidad. Esto no es posible, así que tengo una lista con varios morfemas que quisiera analizar (...).Tengo varias en mente, pero es más la metodología que he desarrollado para analizar una base de datos muy amplia. (...)

C: Sí, es nada más indicativo, no puede ser exhaustivo. (...)

N: **¿Ud. cree que con estos mapas luego se podría ir más a detalle? ¿O piensa que estos mapas ayudan a tener una sobrevista y luego a base de estos mapas encontrar áreas de investigación donde se podría ir más a detalle a estudiar estos topónimos?** (...)

C: Si, yo creo que sí. **Desde luego los mapas son útiles para hacer esta comparación. Pero sí yo creo que sí hay que usar los mapas más indicados.** (...)

N: **¿Ud. cree que esta metodología que Ud. ve aquí con estos mapas se podría aplicar a otros morfemas? Y quizás, si se pudiera aplicar a otros morfemas, ¿qué limitaciones o que adaptaciones ve Ud. que se tendrían que tomar en cuenta?**

C: **Yo creo que es una visión básica pero sí limitada en el sentido de que necesita más análisis, más profundidad en términos históricos, en términos antropológicos. Es un acercamiento que nos puede dar indicaciones para profundizar después. Pero por lo pronto yo creo que sí es un buen avance esta metodología. Yo creo que si está muy bien.**

N: **¿Y Ud. cree que, como historiador o que quizás lingüista, podría reproducir estos mapas fácilmente o cree que un cartógrafo o una cartógrafa debería participar en esta creación de estos mapas o debería tener entre sus objetivos de investigación de crear estos mapas?**

C: Si yo creo que si es útil. Sin embargo, si me gustaría que te comunicaras con Karin Lefebvre, la arqueóloga del Centro de Investigaciones en Geografía Ambiental de la UNAM, porque como te digo ella está en un centro de investigaciones de geografía ambiental y es la materia de trabajo cotidiano de ella los mapas de información geográfica de primer nivel. (...)

N: **¿Cree Ud. que los resultados de este proyecto son significativos para la investigación histórica o lingüista o de otras perspectivas que hay sobre la toponimia?**

C: **Si, si desde luego que sí. Yo creo que sí es interesante tu proyecto, tu avance. Nos presenta una base de investigación a futuro. Pero yo creo que sí nos permite ver la complejidad del tema de la toponimia y es útil para posteriores investigaciones. Yo creo que sí, es muy útil.**

N: Ya lo ha dicho, pero quizás nuevamente: **¿Qué limitaciones ve en estos mapas?**

C: Como te digo **tomar en cuenta otros factores que intervienen en el cambio toponímico, en los aspectos históricos, en las migraciones mismas, la perdida de la lengua, otros factores que hay que tomar en cuenta para un análisis más detallado y que puedas señalar que se requiere tener una etapa posterior de análisis más detallado**. (...)

7.2. Georeferencing Reports

7.2.1. Historical Map (whole extent)

Summary:

Transformation	2nd Order Polynomial
Control Points	28 / 28
Total RMS Errors (Forward, Inverse, Forward-inverse)	22,680.66, 25.88, 1.53

Control Points:

Link	Enabled	Source X	Source Y	Map X	Map Y	Residual	Residual X	Residual Y
1	Y	119.642466	-410.161154	1,177,365.020385	2,062,304.739260	13,512.714060	-13,508.920250	-320.179519
2	Y	684.911996	-1,362.982979	1,715,350.013670	1,246,823.185465	11,759.078888	11,569.534777	-2,102.807963
3	Y	836.659590	-1,050.585930	1,852,908.552844	1,516,286.265495	24,657.150967	23,680.736111	-6,870.067764
4	Y	2,100.256626	-1,075.627988	2,948,676.509004	1,541,543.834178	20,622.214388	19,891.763032	5,439.989868
5	Y	1,984.427835	-1,926.217949	2,897,955.812121	787,255.809232	38,933.751898	31,354.600394	-23,080.859408
6	Y	1,861.782976	-2,018.020521	2,752,358.276139	731,460.294760	9,721.534109	-8,810.622231	4,108.669046
7	Y	1,873.150638	-2,145.973594	2,764,899.364300	619,749.567290	10,609.388244	-10,562.067128	1,000.928022
8	Y	1,796.934225	-2,215.921812	2,725,867.821859	544,064.038292	20,065.369155	16,211.116998	-11,824.496819
9	Y	2,198.263413	-2,205.803461	3,062,291.979550	573,585.357194	9,218.747236	-4,654.037745	-7,957.715330
10	Y	2,077.631834	-2,177.591445	2,960,215.580232	594,071.963735	6,529.281411	1,665.812252	-6,313.207211
11	Y	2,503.274138	-2,323.129828	3,384,054.542619	481,058.807346	43,293.629308	40,811.605786	-14,448.223829
12	Y	2,687.281962	-2,012.691190	3,485,347.640364	758,950.304475	9,117.686886	-6,653.983463	-6,233.515719
13	Y	2,783.473088	-2,189.205289	3,563,143.464126	650,529.722781	37,642.582356	-22,970.346390	29,821.589380
14	Y	3,038.699640	-1,685.573938	3,780,871.096403	1,050,706.337607	2,959.073240	-1,202.895015	-2,703.545454
15	Y	2,940.663551	-1,830.878222	3,697,574.911954	918,560.790884	12,101.837219	-7,558.882424	-9,450.807403
16	Y	2,442.889810	-2,086.291419	3,286,967.497709	687,615.889907	9,409.782503	7,689.244813	-5,423.976489
17	Y	2,113.072141	-1,709.440823	2,975,972.220033	1,000,906.849319	4,991.201327	4,725.553715	1,606.621539
18	Y	1,693.835588	-1,695.002381	2,576,798.871975	1,000,871.141599	24,995.514379	-24,617.785068	4,329.018066

Link	Enabled	Source X	Source Y	Map X	Map Y	Residual	Residual X	Residual Y
19	Y	1,742.621875	-1,127.050644	2,600,451.748939	1,485,662.850376	20,782.212135	-20,511.497572	3,343.472533
20	Y	1,323.455979	-1,727.554924	2,287,809.807251	950,149.949392	13,508.863829	12,386.406518	-5,391.320388
21	Y	1,212.332944	-1,615.024355	2,202,187.362685	1,056,776.039651	29,180.705011	27,937.842147	8,425.587288
22	Y	1,583.094156	-1,965.732293	2,463,528.305536	787,573.312008	55,356.770074	-48,588.159209	26,524.757826
23	Y	1,658.894469	-1,889.476351	2,540,831.460156	835,015.277697	36,691.934724	-36,222.437124	5,850.907845
24	Y	1,781.252208	-1,901.546906	2,673,940.099291	827,120.625572	12,424.165554	-11,890.798576	3,601.221865
25	Y	2,167.247216	-1,890.308284	3,032,493.279128	846,790.303910	5,773.458068	5,662.673109	-1,125.589320
26	Y	2,148.282643	-1,976.502727	3,014,740.861855	769,820.894735	4,376.703553	1,117.684822	-4,231.585356
27	Y	2,256.060788	-1,932.232096	3,115,375.156237	811,071.628819	9,540.628883	8,262.533306	-4,770.130277
28	Y	2,618.900620	-2,275.680759	3,448,882.566409	558,884.136640	18,814.010662	4,785.324416	18,195.264973

7.2.2. Historical Map (subset of Central Mexico)

Summary:

Transformation	2nd Order Polynomial
Control Points	16 / 16
Total RMS Errors (Forward, Inverse, Forward-inverse)	11,557.02, 18.11, 2.21

Control points:

Link	Enabled	Source X	Source Y	Map X	Map Y	Residual	Residual X	Residual Y
1	Y	119.642466	-410.161154	1,177,365.020385	2,062,304.739260	13,512.714060	-13,508.920250	-320.179519
2	Y	684.911996	-1,362.982979	1,715,350.013670	1,246,823.185465	11,759.078888	11,569.534777	-2,102.807963
3	Y	836.659590	-1,050.585930	1,852,908.552844	1,516,286.265495	24,657.150967	23,680.736111	-6,870.067764
1	Y	97.312940	-437.598315	2,723,692.246068	542,486.972065	1,133.204317	816.891051	-785.392281
2	Y	753.657567	-656.191375	3,089,783.603085	419,671.326804	10,223.356810	-8,261.322129	-6,022.257146
3	Y	1,138.533160	-575.527538	3,383,486.962109	480,927.208085	22,854.378099	21,659.336247	-7,293.541778
4	Y	813.369732	-12.180625	3,132,761.488852	797,058.723051	3,361.930247	-3,353.599625	236.525972
5	Y	669.559019	-396.355419	3,061,440.131617	574,038.499141	1,912.488082	-1,540.246582	-1,133.689168

Link	Enabled	Source X	Source Y	Map X	Map Y	Residual	Residual X	Residual Y
6	Y	378.684894	-473.544189	2,883,863.597962	526,984.043125	2,956.395847	1,751.847704	2,381.450405
7	Y	647.602145	-349.058873	3,043,503.645455	598,182.404707	7,966.245169	-7,110.366117	-3,592.179809
8	Y	390.476917	-6.073197	2,901,857.255370	789,314.194194	10,912.241794	10,910.580903	190.382053
9	Y	216.892886	-34.230017	2,790,519.026174	772,514.172445	7,331.541302	-7,244.471612	1,126.556222
10	Y	169.961346	-303.546588	2,765,396.107110	620,626.474978	7,346.109449	-7,333.295610	433.704419
11	Y	484.178616	-353.250192	2,959,904.485657	593,990.918653	8,285.726784	8,118.337415	-1,657.065465
12	Y	549.619560	-347.654673	2,995,627.582418	596,593.419428	5,767.123187	4,499.963514	-3,606.943059
13	Y	602.227768	-544.313000	3,025,809.856385	493,362.988952	14,611.221172	13,403.084735	5,817.654485
14	Y	988.900771	-538.924522	3,246,965.118024	513,052.081457	18,390.882991	-15,490.704523	9,912.751918
15	Y	962.525457	-554.860721	3,223,601.573754	498,201.569512	21,143.593872	-20,420.980280	5,480.431210
16	Y	859.122689	-548.289700	3,185,860.312436	491,232.082372	9,709.699615	9,594.944911	-1,488.387978