University of Zurich[UZH]

Developing a User-Centered Interactive Web Interface for the Glottography Spatial Language Database: A Case Study Focused on South America

GEO 511 Master's Thesis

Author
Silvan Caduff
16-710-899

Supervised by
Prof. Dr. Robert Weibel
Dr. Peter Ranacher
Dr. Tumasch Reichenbacher

Faculty representative
Prof. Dr. Robert Weibel

30.04.2024
Department of Geography, University of Zurich

Abstract

The application of interactive maps and visualizations in academic research has significantly enhanced the analysis of spatial phenomena. However, the field of linguistic geography has not fully benefited from these technological advancements. Existing geovisualization tools lack the specialized capabilities required for mapping language distributions with the necessary granularity and specificity. This gap is particularly acute in the absence of comprehensive databases that document language distributions globally while providing interfaces tailored to the diverse needs of users from various academic disciplines. Addressing this deficiency, this thesis introduces a geovisualization platform developed within the Glottography Project, specifically designed to enhance the exploration and analysis of spatial linguistic data.

The proposed platform provides a polygon-based representation of language areas, enriched with reliable source citations, and designed with a keen focus on the requirements of linguists, geographers, and general scholars. It integrates principles from cartography, human-computer interaction, and information visualization to create a user-centric design that not only supports rigorous academic inquiry but also promotes an engaging and intuitive user experience. By doing so, it aims to bridge the existing gap in linguistic geovisualization, offering a tool that enhances the understanding of linguistic diversity and distribution.

Furthermore, this thesis explores the theoretical underpinnings and practical challenges involved in developing such a platform, highlighting the essential role of a user-centered design (UCD) approach. This approach is critical for ensuring that the platform adequately addresses the diverse functional needs and cultural contexts of its users. Although the platform shows promise, it is understood that it is still in the developmental phase and not yet a finalized solution. The expected outcomes include enhanced data accessibility and improved user engagement with geospatial representations of linguistic data. This study contributes to the ongoing academic discourse on the spatial representation of linguistic data and aims to provide insights and methodologies that could inform future developments in linguistic geovisualization.

Keywords: #Spatial Linguistics, #User-Centered Design, #Glottography, #Interactive Maps, #Usability

Acknowledgments

I extend my deepest gratitude to Prof. Dr. Robert Weibel, Dr. Peter Ranacher, and Dr. Tumasch Reichenbacher for their unwavering support during some challenging times. Their invaluable advice and encouragement have been helpful throughout the course of this thesis.

I am also immensely thankful to Prof. Dr. Paul Widmer, Prof. Dr. Johannes Kabatek, Prof. Dr. Carlota de Benito Moreno, Dr. David Inman, Dr. Erika Just, Dr. Kellen Parker Van Dam, Dr. Sietze Norder, Dr. Tamara Bodden, Meeli Roose, Michael Niederberger, and Geneviève Bénédicte F. E. Hannes for their participation in my user and usability assessments. Their insights and feedback have significantly enriched this work.

Special thanks to Dr. Annina Michel for her review of my interview guidelines, ensuring their effectiveness and clarity.

I am grateful to Gina Schlatter for her proofreading and review of my thesis, which greatly enhanced the clarity and quality of the final manuscript.

This thesis also benefited from the foundational support provided by the Department of Geography at the University of Zurich. The Glottography project, a major component of my thesis, received partial funding from the UZH Research Priority Program 'Language and Space' and the NCCR Evolving Language. This funding was instrumental in bridging the critical gap in spatial language data availability and has profoundly shaped the scope and direction of my research. I deeply appreciate this support, which has significantly contributed to the development of my thesis.

Table of Contents

1.Introduction

The application of interactive maps and visualizations in academic research has proven to be a valuable tool for data exploration (Edsall et al., 2008), yet the field of linguistic geography has seen limited benefit from these advancements (Luebbering, 2011). The particular demands of mapping language distributions call for specialized tools that are currently lacking, underscoring a pressing need in the discipline. This gap is evident in the lack of comprehensive databases that not only document language distributions globally but also provide interfaces that cater to the specific needs of users from various disciplines (Rantanen et al., 2022). Creating satisfactory designs for interactive maps call for a wide range of interdisciplinary knowledge transfer considering the complexity of data visualization (Robinson et al., 2017). Cartography, Human-Computer Interaction (HCI), and information visualization are a small extendable collection of disciplines that can and should be considered when creating map designs with aim of transferability for new use cases (Çöltekin et al., 2017; Griffin et al., 2017).

Geovisualization shares core interests with all of these disciplines but focuses most on geospatial data exploration enabled by digital interactivity, which makes it distinct from the traditional cartography (Çöltekin et al., 2018). This key distinction drives it to be the primary field of study for research with prioritization on geospatial representation, visualization-computation integration, interface design and usability issues (Dykes et al., 2005). Geovisualization, with its focus on geospatial data exploration through digital interactivity, provides the theoretical and practical framework necessary for the development of effective visualization platforms. This integration of geovisualization principles enriches the toolset available for creating platforms that can adeptly handle the complexities of mapping language distributions.

Few global linguistic databases exist with spatial information. There is the Ethnologue (Eberhard et al., 2023) a pay-walled database documenting language distribution in the world, containing polygon data, but there are no academic references to back up the correctness of the languages ranges. Another project, Glottolog (Hammarström et al., 2022), initially started as an academic bibliography about the world's languages and their classifications. Its geospatial data is primarily limited to point data for most languages and lacks speaker counts or ranges. One last example would be the Native Land [1] that aims to share knowledge on indigenous territories, treaties, and languages mainly focused on the Americas, Australia and sporadically in Africa. Native Land uses polygonal data to represent language ranges which can be interactively accessed and it reveals the sources to the spatial information. It is limited to its spatial coverage and as its name implies, has a focus on indigenous languages.

Rantanen et al. (2022) have recently made efforts to provide guidelines on how to put qualitative varying language data into an effective and standardized data structure. Furthermore, they documented their process of developing the Uralic language distribution database and used the preprocessed language data to incorporate it into an interactive web map. It emphasizes the importance of harmonizing spatial language data is crucial to develop a robust visualization. Moreover, the project points to frequently addressed problems in both Geographical Information Science and DataVis such as polygonal overlap and multidimensional data visualization. Given these considerations the project naturally leads to common

[1] https://native-land.ca/

questions. These include inquiries about the representation of multilingual areas, handling of multiple sources for a single language and the integration of the hierarchical data nature of languages (Luebbering, Kolivras, and Prisley 2013; Rabanus 2019; Hoch and Hayes 2010). This makes the Uralic Historical Atlas project (see Section 2.5.) a valuable example of how to gather and visualize information on an entire language family for a map interface. However, the atlas is limited by its linguistic and geographical extent and further studies with special focus for global language mapping in web interfaces are needed.

Due to the dissatisfying state of options, the Glottography project has been initiated by the GIS Unit at University of Zurich. It aims to fill in gaps of the established databases by providing an open-source web atlas focusing on language distribution using source referenced polygons. The Glottography has started collecting and standardizing data and is currently developing a graphical user interface for linguists, geographers and general scholars with interest. The project seeks an academic approach for designing the map interface and a User-Centered Design (UCD) study could achieve the desired evaluation process of how the data exploration can be optimized for researchers.

It's becoming clear that integrating user insights into the design process is essential (Dykes et al., 2007; Nakić et al., 2022). This need arises from the recognition that developers may not always have comprehensive knowledge of the specific requirements and experiences of end-users. UCD addresses this gap by involving users directly in the design process, utilizing their domain expertise to shape solutions that are both functional and user-friendly (Mao et al., 2005; Vredenburg et al., 2002). UCD is particularly important in fields like geospatial language visualization, where user requirements can vary widely and where the application must adapt to diverse linguistic and cultural contexts. However, there is still a noticeable lack of user-centered studies in the design and evaluation of interactive maps, especially those dealing with spatial language data (Roth et al., 2017). The existing literature often focuses on outdated static mapping techniques, failing to capture the dynamic nature of today's geospatial challenges (Haynie and Gavin, 2019).

There is a clear research gap in spatial language visualization for map interfaces, and the Glottography project needs to develop a user-oriented platform for its linguistic database. This situation strongly supports the need for an in-depth evaluation to address and fill this research gap. The decision to employ a user-centered case study in the Glottography aims to address the research gap in spatial language visualization for map interfaces. This UCD approach, highlighted by Roth (2019) for its ability to uncover hidden requirements and design issues, aligns with the project's needs. However, it's important to note that the success of this methodology will be evaluated through the actual development of the project's interface. Engaging users throughout the design and evaluation process is expected to improve the interface’s relevance and functionality. This involvement is hypothesized to yield a platform that effectively meets user expectations and contributes to the development of design guidelines that could be beneficial for similar initiatives. The validation of this user-centered approach will be determined upon the project's completion.

2. Related Work

2.1 Language: Definitions and Dynamics

Language is a vital communication tool that enables individuals to express thoughts, ideas, and emotions. Each language possesses unique rules for constructing words and sentences, setting it apart from others. More than a mere tool for daily communication, language also embodies the cultural heritage and identity of its speakers, preserving traditions, knowledge, and history (Deutscher, 2006).

Languages continuously evolve, shaped by societal changes, technological advancements, and intercultural exchanges. They are grouped into families that share common origins and isolates that stand apart without clear relationships to other known languages. This classification encompasses various languoids, which include language families, individual languages, and dialects. The term languoid refers to any group of lects or languages assembled for specific analytical purposes (Good and Hendryx-Parker, 2006). This grouping can include straightforward categories like individual languages or dialects, as well as more complex assemblies such as genealogical or areal groups of languages. Isolates are of particular interest because of their distinctive linguistic characteristics and their isolation from other languoids.

The Glottolog database (see Section 3.1.) is instrumental in this classification process, employing historical-comparative research to systematically assign a unique glottocode (Forkel and Hammarström, 2022) to each languoid for precise identification. This method involves comparing linguistic features across languages to establish their historical relationships, thereby mapping out a genealogical tree (see Section 3.2.4.) that outlines how languages diverge and evolve over time. The Glottolog also maintains 'bookkeeping' languoids for entities not recognized as distinct but included for completeness. With 7,665 primary languages classified into 246 families and 184 isolates, the Glottolog provides an essential framework for the academic study of linguistic diversity (Campbell and Poser, 2008; Evans, 2009).

These efforts are vital for understanding the dynamics of language development and for addressing challenges like multilingual area representation, multiple source integration, and the hierarchical nature of languages (Luebbering, Kolivras, and Prisley 2013; Rabanus 2019; Hoch and Hayes 2010). Thus, the ongoing research and data refinement in the Glottolog are crucial for documenting global language distributions and supporting comprehensive linguistic studies.

2.2 The Glottography Project

The Glottography, initiated by the Department of Geography at the University of Zurich with partial funding from the UZH Research Priority Program 'Language and Space' [2] and the NCCR Evolving Language[3] directly addresses a critical gap in spatial language data availability. This initiative is strategically designed to develop a comprehensive, publicly accessible database that focuses on the geographical distribution of languages worldwide.

[2] https://www.spur.uzh.ch/en.html
[3] https://evolvinglanguage.ch

Central to the Glottography is the creation of an open-access spatial language database that utilizes polygonal language ranges. The ongoing development of a user-friendly interface, which is the focus of this thesis, is crucial for the project's success. Once completed, it will enable researchers, educators, and the general public to easily explore and interact with the data, significantly enhancing the understanding of global language distribution. By addressing the methodological limitations historically associated with linguistic research, which has often concentrated on non-spatial aspects of language (Haynie, 2014; Rantanen et al., 2022), the Glottography embodies a shift towards integrating spatial data into the study of human history.

The Glottography not only meets the growing demand for a robust and accessible linguistic database but also sets a new standard for global language mapping initiatives. It underscores the importance of a methodical approach to geovisualization, such as User-Centered Design, ensuring that the complexities of global language mapping are comprehensively addressed. This project is poised to significantly advance the field of geolinguistics by providing a model that can be emulated worldwide, enhancing the potential of interactive web interfaces in linguistics research.

2.3 Significance of Spatial Language Data in South America?

This research explores the challenges and potential solutions for visualizing language distribution. South America serves as a well-suited testing ground for this investigation due to the characteristics of its spatial language data within the Glottography. While the Glottography's overall data collection is not extensive, the South American portion offers a sufficient volume for a focused case study. The complexities encountered in visualizing these languages mirror the intricacies typically found in language distribution across the globe (Rabanus, 2019). Examining these challenges, which will be further discussed in Section 3.3., within the South American context allows for the development of broadly applicable solutions for future projects in global language visualization.

South America has a remarkable diversity of indigenous languages, making it a unique and fascinating region for studying language visualization (Van Gijn et al., 2017). Estimates suggest the presence of 108 documented language families, with a significant portion comprising smaller language groups (Campbell, 2012; Hammarström et al., 2022). This linguistic richness necessitates effective strategies for visualization and analysis to ensure the survival of this vital cultural heritage. Language is intricately linked with the culture and identity of indigenous communities across the continent, and preserving this linguistic diversity is crucial (Campbell, 2012). Each language offers a unique window into the history, culture, and knowledge of its speakers. For instance, the Quechua language, spoken by millions in the Andean region, carries a rich oral tradition of storytelling and poetry that reflects the worldview and history of these communities (Campbell and Grondona, 2012).

Spatial language data plays a vital role in language preservation efforts. By identifying areas rich in linguistic diversity, such data allows for prioritization of documentation and revitalization projects (Lee et al., 2022; McCarty and Coronel-Molina, 2016). In South America, where many indigenous languages face threats from dominant language spread, globalization, and speaker loss (Crevels, 2012; Evans, 2009), mapping language distribution empowers researchers and policymakers to develop targeted strategies for linguistic and cultural sustainability. Interactive maps and atlases facilitate the visualization of spatial language data, making this information accessible to a broader audience, including educators, policymakers, and the general public (Upton, 2011). This accessibility fosters greater awareness and appreciation of linguistic diversity, potentially leading to increased support for language preservation initiatives, like the Native Land Digital.

While colonial languages like Spanish and Portuguese play significant roles in the continent's linguistic heritage (Severo, 2016; Severo and Makoni, 2014; Stavans, 2021), the focus of this thesis is on the indigenous languages. Notably, in countries like Bolivia and Paraguay, a substantial portion of the population communicates in indigenous languages like Aymara, Quechua, and Guaraní, highlighting their prevalence (Hamel and Butragueño, 2004). These languages play a crucial role in the multilingualism of the region, often coexisting with Spanish or Portuguese. The challenge lies in the dispersion and diversity of these indigenous languages, which, unlike the widespread colonial languages, are scattered across various locations and represent a wide range of linguistic groups. This dispersion presents unique challenges for linguistic documentation and analysis. Despite the historical focus of linguistic atlases on colonial languages, recent scholarship is increasingly recognizing and documenting indigenous languages, marking a significant shift towards inclusivity in linguistic research (Rabanus, 2019).

2.4 Geovisualization for Language Representation

Geovisualization employs interactive maps to visualize spatial information. The shift from static to dynamic, interactive language maps significantly enhances how we engage with and understand linguistic diversity. Unlike traditional printed maps, which are limited by their inability to change scale, add detail, or link to multimedia resources, digital platforms offer a real-time, interactive exploration of languages (Dykes et al., 2005). Users can navigate through data, customize views for specific languages, and conduct side-by-side language comparisons. This advancement addresses previously identified limitations, such as interaction constraints and the difficulty of incorporating enriching elements like sound, pictures, or video into linguistic data presentation (Cartwright and Peterson, 2007).

Online linguistic resources such as The World Atlas of Language Structures (WALS) (Dryer and Haspelmath, 2013), Ethnologue (Eberhard et al., 2023), and Glottolog (Hammarström et al., 2022) play a crucial role in the digital representation of languages. WALS primarily concentrates on linguistic properties and features rather than individual languages, offering a unique perspective on linguistic diversity. Ethnologue offers extensive spatial information on language distribution, including language branches and families, but access to its full database is restricted by a paywall, and the methodologies employed are not transparently detailed. Glottolog, while offering valuable data on language locations, predominantly provides point data, which may not fully capture the geographic extent of language territories as elaborated in the subsequent paragraph. Despite the valuable insights offered by these databases, there is a significant gap in their coverage of linguistic geography, particularly in representing language areas with polygonal data. Such data are essential for analyzing language territories and understanding potential overlaps among languages. This limitation underscores a critical research gap: the need for comprehensive geographic representation of language speakers and communities. Recent research (Rantanen et al., 2022) emphasizes absence of detailed polygonal data on language distribution what limits the utility of these databases for in-depth linguistic research, pointing to an essential area for future development.

The absence of web-map services offering free access to polygonal representations of language areas underscores a significant gap in the field of linguistic geography. This limitation hinders detailed analysis of linguistic diversity and its geographic distribution (Luebbering et al., 2013). Addressing this challenge requires the consolidation of spatial language data into a digital format that supports the visualization of language territories. Efforts to collect and standardize spatial language data (Forkel et al., 2018; Rantanen et al., 2022) are crucial in bridging this gap, targeting the inconsistent availability and quality of spatial information. Much of the data on language distribution currently exists in analog forms, such as maps and textual sources, necessitating a transition to digital for effective utilization in new mapping visualizations or for integration with other spatial datasets. The process of digitizing and standardizing these data into an unified spatial database is essential for improving accessibility and facilitating linguistic research, ultimately contributing to a comprehensive global database that reflects the world's linguistic diversity (Rantanen et al., 2022).

2.4.1 Point vs. Polygonal Data: Differences in Language Representation

Existing methods for representing language distribution often rely on point data, pinpointing specific locations where a language is spoken (McNew et al., 2018). While this approach offers a basic presence indicator, it fails to capture the intricacies of language spread and the complexities of language territories. Point data struggles to represent the nuances of language dominance within a region or areas of bilingualism (Luebbering, 2011). However, developing a standardized mapping methodology for language polygons presents challenges (Rantanen et al., 2022). Although polygonal data provides detailed geographic outlines, language boundaries are inherently fluid, influenced by migrations, cultural exchanges, and social dynamics (Daurio and Turin, 2019). The creation of a standardized approach would significantly ease research by ensuring a consistent quality standard, but this is beyond the scope of this thesis and remains a crucial future consideration for the field.

To enhance uniformity and comparability across different studies, polygon data is favored despite the fluid nature of language boundaries. This preference arises because, while language borders are not fixed, polygons offer a more precise and comprehensive visualization of linguistic areas than point data. They better represent the extent of language usage and can adapt to show the degrees of language dominance and transition zones, crucial for advanced linguistic and geographical analyses.

Nevertheless, in scenarios where spatial information needs to be presented at a more general level or the occurrence of language is point-like (Rantanen et al., 2022), such as in a single village, the use of point data can be equally valid. Point data simplifies representation when the detailed geographic extent of a language is not required or is impractical. This thesis focuses on polygonal data due to its broader applicability in representing linguistic distributions. It is essential to contextualize the choice of data type in linguistic mapping, ensuring alignment with research objectives and the specific nature of the language distribution under study.

2.5 Uralic Historical Atlas and Native Land Digital

The Uralic Historical Atlas, finalized in 2024, serves as an exemplary case within the broader context of this thesis, which addresses the significant gaps in current geovisualization tools for studying language distribution. The URHIA, with its detailed, polygon-based depiction of the Uralic language family, embodies the principles and objectives that the thesis proposes for the Glottography, a platform designed to enhance the exploration of spatial linguistic data.

The URHIA initiative is noted for its inclusive and participatory design process, closely aligning with the UCD principles emphasized in my thesis. By involving a wide range of stakeholders, including linguists, historians, GIS experts, and IT professionals from the project's inception, URHIA ensured that the final product was not only scientifically rigorous but also tailored to meet the specific needs of its users (Roose et al., 2021). This approach mirrors the methodology employed in the Glottography, where interview participants were selected across a spectrum of expertise, encompassing linguistics, history, GIS, and IT, to ensure diverse input that would improve the interaction and usability studies as well as the conceptual development and prototyping stages.

In developing URHIA, significant emphasis was placed on initial core group meetings, pivotal in defining the platform's purpose and scope. These meetings facilitated the integration of diverse expertise into the atlas, enhancing its depth and relevance. Unlike URHIA, which utilized group meetings and workshops, the Glottography adopted a different approach by conducting individual qualitative interviews to gather insights (see Section 5.2.). This method allowed for in-depth, focused feedback on functionalities, branding, and design, aligning with the interaction and usability studies emphasized in this thesis. This shift in methodology is critical for ensuring that the platform not only serves academic purposes but is also intuitive and engaging for users, adapting the UCD process to a more personalized interaction framework.

The URHIA project's development utilized a framework that classified key functionalities of spatial data platforms into five categories: platform type, layer management, spatial data query, cartographic map elements, and presentation of attributes (see Chapter 5.). This framework, derived from a thorough review by Roose et al. (2021) of established platforms, informed the incorporation of essential features into URHIA. By identifying advanced functionalities such as interactive mapping, transparency control, and detailed attribute presentation, the project team was able to create a platform that is not only user-centric but also rich in analytical capabilities. This systematic approach to integrating crucial features guided the URHIA's design to meet the specific needs of linguistic geovisualization, offering a model that my thesis could consider for the Glottography to enhance the exploration and presentation of linguistic data.

Local initiatives like URHIA demonstrate the feasibility and value of presenting language areas as polygons within a scientifically rigorous framework. Similarly, Native Land Digital focuses on mapping Indigenous languages and cultural territories, driven by user engagement and support. Both projects underscore the potential of specialized linguistic databases to deeply engage audiences and provide rich, contextualized data visualization. Their success in fostering a greater appreciation for linguistic and cultural diversity offers key implications for global linguistic documentation and preservation.

Building on this foundation, the Glottography aims to extend these concepts to a wider canvas, seeking to cover the global spectrum of languages. Unlike its predecessors, Glottography is designed to encompass all languages, integrating the detailed geographic and cultural insights characteristic of URHIA and Native Land Digital into a comprehensive, universally accessible platform. This approach not only amplifies the depth and breadth of linguistic data available but also expands the potential for research and public engagement across the entire linguistic landscape.

2.6 Introduction to User-Centered Design

Building on the foundation laid by projects like URHIA and Native Land Digital, which have successfully utilized specialized databases to engage users and visualize linguistic data, the next step involves creating a platform that not only broadens the scope of these projects but also ensures their accessibility and usability. This transition towards a more inclusive and user-friendly platform calls for a methodology that places the user's needs and experiences at the forefront of development, User-Centered Design (UCD).

UCD focuses on involving users throughout the design and development process, ensuring that the final product is accessible, easy to navigate, and meets the real-world needs of its audience (Norman and Draper, 1986). This method emphasizes integrating user feedback throughout the design process, ensuring the product is both functional and user-friendly, thereby addressing specific user needs (Roth, 2019). UCD has proven helpful for enhancing application usability (Hasani et al., 2020; Mao et al., 2001; Vredenburg et al., 2002).

For the Glottography, which aims to offer global coverage of languages with an emphasis on detailed spatial data, adopting a UCD methodology means prioritizing the usability and accessibility of the platform for linguistic researchers and enthusiasts. By aligning the development process with UCD principles, the project is set to contribute to the ongoing discussion about effective map design and usability. Further exploration of UCD principles in the subsequent section will demonstrate their critical role in the development of the Glottography.

2.6.1 User-Centered Design in Geovisualization for Linguistic Data

User-Centered Design (UCD) is a framework that emphasizes the importance of user needs and experiences throughout the design and development process. It ensures that digital tools are accessible, efficient, and tailored to diverse user preferences, which is essential for enhancing user engagement and comprehension in complex domains. Unlike prescriptive models, UCD operates as a dynamic, multi-step process in which interactive maps and tools undergo continuous evaluations against established criteria, resulting in iterative refinements and improvements based on identified deficiencies (Marsh and Haklay, 2010). Integrating UCD into the design and evaluation of interactive systems is crucial for continuously improving user interaction (Roth et al., 2017).

UCD follows an iterative process that cycles through design, feedback, and refinement, ensuring the product meets user needs effectively. This approach includes documenting development, integrating user feedback into feature implementation, and conducting evaluations through methods like task-based scenarios and usability testing (e.g., SUS). Outlining UCD's high-level overview early provides a foundation for understanding its application in creating accessible geovisualization platforms for linguistic data.

In geovisualization, UCD plays a pivotal role in presenting geospatial data interactively and accessibly (Tsou, 2011). It leverages insights from cartography, human-computer interaction (HCI), and information visualization, creating tools that are intuitive and align with user expectations for navigating spatial information. This interdisciplinary effort makes geovisualization tools more responsive to user needs, improving the usability of digital platforms (Çöltekin et al., 2017; Griffin et al., 2017). Geovisualization, in particular, distinguishes itself by prioritizing the exploration of geospatial data through digital interactivity, addressing both the representation of languages and the usability of the platform (Dykes et al., 2005).

Incorporating UCD into the development of a geovisualization platform for linguistic data involves a series of structured activities, as outlined by standards like ISO 9241-210 (International Organization for Standardization, 2019) and ISO 13407 (International Organization for Standardization, 1999). These activities ensure that user feedback is central to the design and development process, aiming to produce a platform that is both useful and resonant with its intended audience. The workflow for completing such a project might encompass a variety of approaches that align with aforementioned UCD principles:

- **To understand and specify the context of use**: Initially, it could involve evaluating the available Glottography's data for completeness and characteristics to fully understand the context in which the platform will be used. This step is crucial for ensuring the data's readiness for visualization and identifying the specific needs it should address.

- **To specify the user and organizational requirements**: Identifying early challenges can inform the process of specifying user and organizational requirements. Including as much relevant information as possible, for instance, from databases like Glottolog, might maximize the utility and relevance of the platform.

- **To produce design solutions**: Developing a basic prototype could serve as a practical step to give stakeholders a first impression of potential functionalities and the scope of data visualization available. This approach allows for the early testing of concepts and the refinement of design solutions based on initial feedback.

- **To evaluate designs against requirements**: Creating qualitative interview guidelines (see Appendix A) to unveil user needs and gathering insights through interviews might further refine the understanding of user requirements. Translating these insights into user stories ensures they are actionable and relevant. Evaluating which user needs are essential for the visual platform allows for a focused development of features that are both functional and requested by users.

2.6.2 Case Studies in User-Centered Design

The effectiveness of User-Centered Design (UCD) methodologies in developing geovisualization platforms is well-documented across a variety of research efforts, highlighting the critical role of user engagement in creating accessible and intuitive interfaces. The work of Kramers (2008) on the Atlas of Canada, as well as Nelson and MacEachren's (2020) research on geovisualization applications for quantified-self data, illustrate the practical benefits of integrating UCD from the beginning of the project. These studies demonstrate the potential for UCD to produce tools that not only meet technical requirements but are also highly attuned to user needs and preferences. Moreover, the design of geovisualization tools for epidemiology underscores the multifaceted advantages of integrating usability techniques within a UCD framework (Robinson et al., 2005). This approach resulted in enhanced data accessibility and user interaction, showcasing the importance of iterative feedback and usability testing in the development process.

The analysis of multidisciplinary spatial data platforms for human-history research (Roose et al., 2021) offers a structured exploration into the application of UCD in managing complex datasets. Their systematic review and pilot study of a UCD process for a spatial data platform reveal essential considerations for data sharing, platform functionality, and the design process. Key to their findings is the emphasis on the necessity of input and expertise from the end-users, especially scholars in human history, and the involvement of geographic expertise to foster a unified understanding across various domains.

These examples affirm the value of employing UCD methodologies in the creation of an explorative linguistic data platform. By prioritizing user feedback, accommodating diverse data types, and engaging in iterative design and evaluation cycles, UCD ensures the development of platforms that are not only technically robust but also deeply resonant with user experiences. This alignment with user needs enhances both the functionality and the user satisfaction of platforms for spatial linguistic data, offering a model for future projects in linguistic geovisualization and beyond.

2.6.3 UCD for Linguistic Visualization

Integrating Nelson and MacEachren's (2020) user-centered evaluation process with the ISO's guidelines for UCD lays a foundational framework for this thesis, particularly in linguistic geovisualization. Their combined methodologies offer a detailed blueprint for implementing UCD principles, providing a structured approach to develop geovisualization platforms. This synthesis not only validates their methods but also ensures that the resulting platforms are well-suited to the needs of linguists and researchers. By adopting and adapting this approach, this thesis aims to contribute to the exploration and preservation of linguistic diversity, justifying the detailed consideration of this work as a central methodological reference.

Firstly, the integration of scenario-based design techniques directly aligns with the ISO activity "to understand and specify the context of use." (as discussed in Section 2.6.1.). Extensive feedback gathered through varied means comprehensively assesses the context in which the platform would operate, ensuring that the design is informed by real-world usage scenarios and challenges. This initial stage lays the groundwork for developing a solution deeply rooted in the users' operational environment.

Secondly, the formulation of hypothetical use case scenarios and conducting a claims analysis to characterize domain problems and design challenges mirror the ISO activity "to specify the user and organizational requirements." (as discussed in Section 2.6.1.). Collaborative efforts define clear user and organizational requirements that guide the subsequent design and development stages, ensuring the evolving design remains tightly coupled with the intended audience's specific needs and preferences.

Thirdly, the insight discovery stage, though primarily focused on summative assessment, contributes significantly to the ISO activity "producing design solutions." (as discussed in Section 2.6.1.). The recruitment of domain and visualization experts to evaluate the platform's support for insight discovery leads to iterative design solutions refinement, informed by expert feedback.

Lastly, evaluating the platform's usability and utility encapsulates the ISO activity "to evaluate designs against requirements." Assessments through semi-constrained tasks and structured surveys confirm the platform's efficacy in meeting predefined user and organizational requirements, highlighting areas for further improvement.

The approach outlined in the user-centered evaluation of cartographic interfaces offers a practical framework that resonates with the structured activities defined by ISO standards for UCD. This methodology effectively demonstrates how these ISO-specified activities can be applied in developing geovisualization platforms, providing valuable insights for integrating linguistic data with user needs in South America.

2.7 Research Gap and Question

The development of geovisualization platforms specifically designed for exploring spatial linguistic data within the Glottography underscores a critical research gap: Currently, there is no freely accessible web map service that adequately visualizes language ranges with the necessary detail. Such a service must represent language areas with polygons rather than mere points, ensure accurate citation of source material, and embody a user-centered design to enhance data exploration. The proposed service aims to provide a platform that is not only rooted in the rich datasets of Glottography but also designed to make this information accessible for exploration by linguists and researchers working within the project. This platform will serve as both a tool for discovery and a reliable scholarly resource, specifically for Glottography's data.

To bridge this research gap, the project will develop a geovisualization platform grounded in User-Centered Design (UCD) principles, focusing on usability, data relevance, and visual appeal. The platform will be engineered to support informed decision-making and encourage research into linguistic diversity. Key to this development is the integration of user feedback, guaranteeing the platform not only displays spatial linguistic data accurately but also meets the actual needs and preferences of its users. The significance of incorporating user feedback into the design process cannot be overstated; it ensures that the platform not only meets the technical requirements of displaying spatial linguistic data but also aligns with the specific needs and preferences of its users.

The project's goal is to design and test a web map service that fills the current void by mapping language distributions in South America with a user-friendly interface. This service will integrate scientifically referenced linguistic data, ensuring that it not only serves as a tool for exploration but also as a reliable resource for academic study. By making linguistic data exploration more intuitive and meaningful, the service aims to support the wider objectives of linguistic preservation and research, making a significant contribution to the accessibility and understanding of linguistic diversity.

Hypothesis: Influence of UCD on the Efficacy of Spatial Linguistic Platform Design

H: Direct user involvement in the development process, coupled with strategic design choices in the platform's interface, emphasizing easy-to-use navigation and clear representation of linguistic data, will significantly improve usability and the effective presentation of spatial linguistic data, leading to high overall user satisfaction.

- **Rationale**: Employing user feedback throughout the development stages and prioritizing user-friendly design elements, such as intuitive navigation and transparent linguistic data visualization, are expected to closely align the platform with user preferences and requirements. This approach aims to simplify the exploration of linguistic diversity, improving user engagement, comprehension of spatial language distributions, and satisfaction.

- **Test**: The evaluation will combine a system usability survey and a task-based assessment post-user interaction with the platform. This survey will measure usability and information relevance, while the task-based evaluation will focus on how specific design decisions - easy navigation and clear data presentation - impact user experience and exploration effectiveness. Through subjective feedback and objective usability metrics, this comprehensive assessment will gauge the platform's success in meeting user expectations and facilitating linguistic data exploration.

3. Data

3.1 Glottolog Database Overview

The Glottolog database serves as a guiding data source in this project. Its purpose is to compile a comprehensive catalog of the world's languages, categorizing them into families and languages. For each language entity—referred to as a languoid (Hammarström et al., 2022)—Glottolog assigns a unique identifier, the glottocode, to facilitate precise tracking and referencing throughout linguistic research.

Table 1 - Glottolog *Dataset Attributes with Description*.

Attributes	Description
glottocode	Unique identifier for each languoid, essential for identification
family	Represents the top-level language group or most common ancestor of the languoid
name	The standardized name of the languoid for easy reference
parent_glottocode	Indicates the direct ancestor within the linguistic family tree, if applicable
color	Hex color value

The dataset extraction was executed using the Python API "pyglottolog"[4], after which the data were rendered to the core attributes listed. The dataset's structure supports both genealogical insights and specific lineage tracing, foundational for the platform's development. Each languoid's glottocode and its corresponding parent_glottocode enable the construction of a hierarchical tree structure. By employing an recursive algorithm, it's possible to navigate through this structure, tracing the lineage and relationships from parent to child languoids. This capability is not only essential for visualizing linguistic connections but also for enabling complex genealogical queries within the platform's scope.

3.2 Spatial Data from the Glottography

The spatial data for this project comes from the Glottography team's efforts to map where languages are spoken across South America. To build this collection, the team asked researchers to share maps with spatial language details. Most of these maps had to be digitized because they were old, or the original digital files were lost. The digitization was done by a GIS master's student, who, while skilled in mapping, had

[4] https://github.com/glottolog/pyglottolog

limited knowledge of linguistics. This might lead to some issues with how languages are labeled and matched with their glottocodes (Forkel and Hammarström, 2022). To make sure the spatial data is accurate, a linguist will review the classifications and glottocode assignments before the project goes live. This step is crucial to ensure the data's accuracy for anyone who will use it for linguistic research or for exploring the platform. The spatial data, when paired with the Glottolog's genealogical information, gives a detailed view of not just who speaks a language, but also where it's spoken, creating a rich resource for users interested in the languages of South America. For this thesis, the focus is specifically on the polygonal dataset, despite the availability of point data. The decision to emphasize language ranges over point representations stems from a desire to go beyond the insights offered by Glottolog's existing point-based data. The challenge lies in developing a platform that effectively showcases these language ranges, providing a more comprehensive and geographically detailed exploration of linguistic diversity.

Table 2 *- Map Dataset Attributes with Description.*

Attributes	Description
id	Primary identifier for the source map
source_id	Connects to the Source dataset, detailing the origin of the map.
location_in_source	Specifies where within the source the map can be found
url	Provides a direct link to the source map if available online
private_uri	An internal reference to the map within the project's database

The Map dataset documents the source maps used to derive the linguistic polygons. It contains detailed metadata for each map, including its origin and how it can be accessed, whether online or in a physical location. This dataset ensures the transparency and reproducibility of the spatial analysis, providing a basis for verifying the accuracy of the polygon data. It supports the research's methodology by cataloging the provenance of the maps, which is crucial for academic rigor and future studies that may wish to extend or replicate this work. These sources are collected from scientific research, which allows for multiple maps to be associated with a single source. Although the current collection represents only a small selection of what is available globally, it provides a cross-section suitable for developing visualization strategies for a platform that could eventually be scaled to a global level. The source information is maintained in a BibTeX format, allowing for straightforward retrieval of detailed bibliographic data related to each map. This connection ensures that the references can be easily located for each source, facilitating academic citation and further research.

Table 3 *- Polygon Dataset Attributes with Description.*

Attributes	Description
id	Primary identifier for the polygon
map_id	Connects to the Map dataset, indicating the source map of the polygon.
original_name	Name of the language as originally listed in the source
glottocode	Unique identifier for each languoid
geometry	Spatial representation of language range
reference_date	Date associated with the language data

The Polygon dataset is designed to provide a spatial representation of language distribution across South America. It converts linguistic data into geographic polygons that define the area where each language is or was spoken. This dataset includes several key attributes, such as the original language name and the reference date for the data, enabling analysis of language spread and changes over time. It is instrumental for studying linguistic diversity and distribution patterns, offering a clear visual method to understand the geographical scope of different languages.

3.3 Data Characteristics

The dataset contains 2,448 polygons that map the spatial extent of languages in South America. Among these, 1,587 polygons have specified ancestral dependencies in order to establish a genealogical route. The dataset encompasses 451 unique glottocodes, which represent the same number of distinct languoids. The Arawakan family has the largest number of languoids represented in the dataset, totaling 60 polygons. Additionally, there are 49 isolates that are represented in polygonal form in this dataset, highlighting areas where languages exist without a known language family. Spatially, the polygons exhibit a considerable range in size. The smallest polygon covers an area of approximately 2.32 square kilometers, while the largest spans about 13,800 square kilometers. The median area size of the polygons is around 2,930 square kilometers, with a mean size of 23,200 square kilometers. This variation indicates a significant diversity in the territorial spread of the languages, from those with limited geographical footprints to others with extensive ranges.

3.4 Data Selection

In the data selection process, priority was given to the completeness of datasets to support the application's functionality. Datasets that were not comprehensive, for instance, those with sparse 'speaker count' information, evident in only a small portion of polygons, were not included in the initial development phase. This exclusion was necessary due to the potential impact of incomplete data on the application's performance and user interface. Additionally, the dataset contained polygons unassigned to any glottocode, referred to as 'bookkeeping'. These polygons represent areas with insufficient or unclear linguistic data that have not yet been fully researched for inclusion. Their inclusion in the dataset ensures spatial comprehensiveness and allows for updates as more linguistic information becomes available, ensuring the dataset remains useful for ongoing research.

3.4.1 Data overview

The Figure 1 illustrates the extent of overlap among language polygons across South America. To create this visualization, the original polygon dataset underwent a process of explosion where overlapping regions were divided into smaller, distinct polygons. These exploded polygons, which are the result of decomposing larger polygons into multiple smaller ones to better distinguish areas of overlap, were then analyzed to determine the number of original polygons that intersect in each area. The shading in the map corresponds to the degree of overlap, with varying intensities of green denoting the count of intersecting polygons.

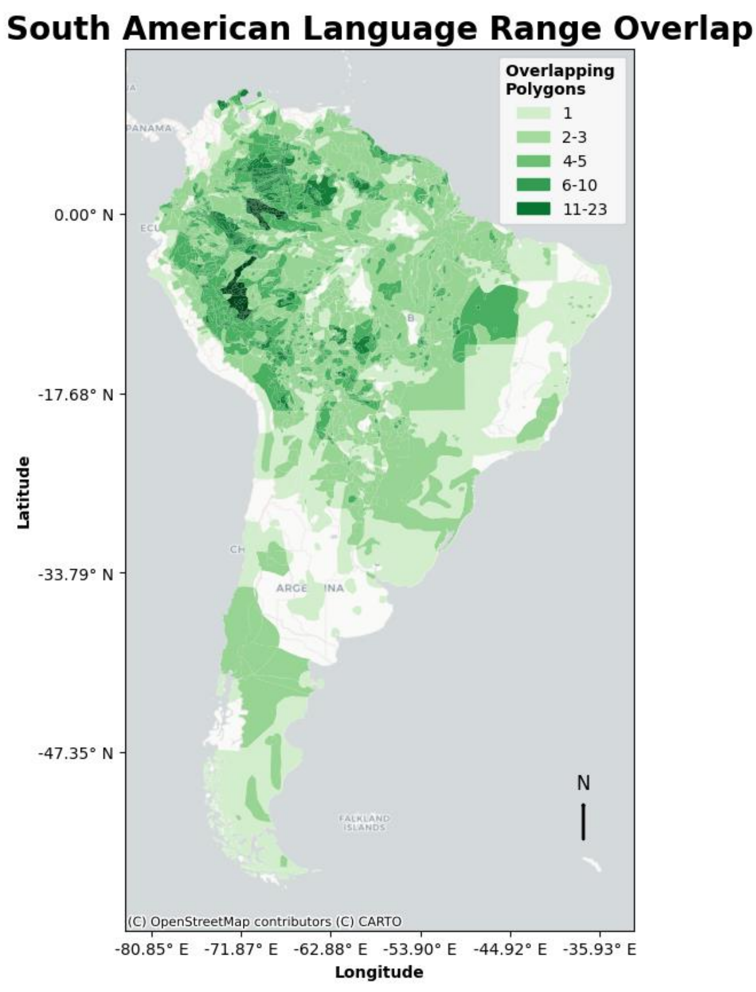


Figure 1 *- Language Polygon Overlap in South America.*

Areas with darker shades reveal a higher overlap, indicating regions where multiple languages have coexisted or still do. Despite the relatively small size of the sample dataset, the presence of such overlaps is noticeable. This demonstrates that even within a limited dataset, there are significant complexities in language distribution. The visualization underscores the necessity for a robust strategy in data representation, as the complexity observed on this smaller scale can be indicative of even greater challenges when dealing with the full dataset. The Figure 1 serves as a key reference point for understanding the potential intricacies involved in visualizing language range overlaps, emphasizing the need for careful consideration in the design and implementation of data representation methods for linguistic geographies.

3.4.2 Language Tree Characteristics

In phylogenetic linguistics, tree diagrams (see Figure 2) are employed to illustrate the relationships between languages, akin to how they are used in biology to show evolutionary relationships between species. In these linguistic trees, nodes represent individual languages or dialects.

The edges, or branches, that connect the nodes represent hypothesized linguistic evolution, indicating how languages may have diverged from common ancestral languages over time. The root of the tree represents the hypothetical proto-language, the ancient ancestor from which the depicted language family has evolved (Campbell, 2007).

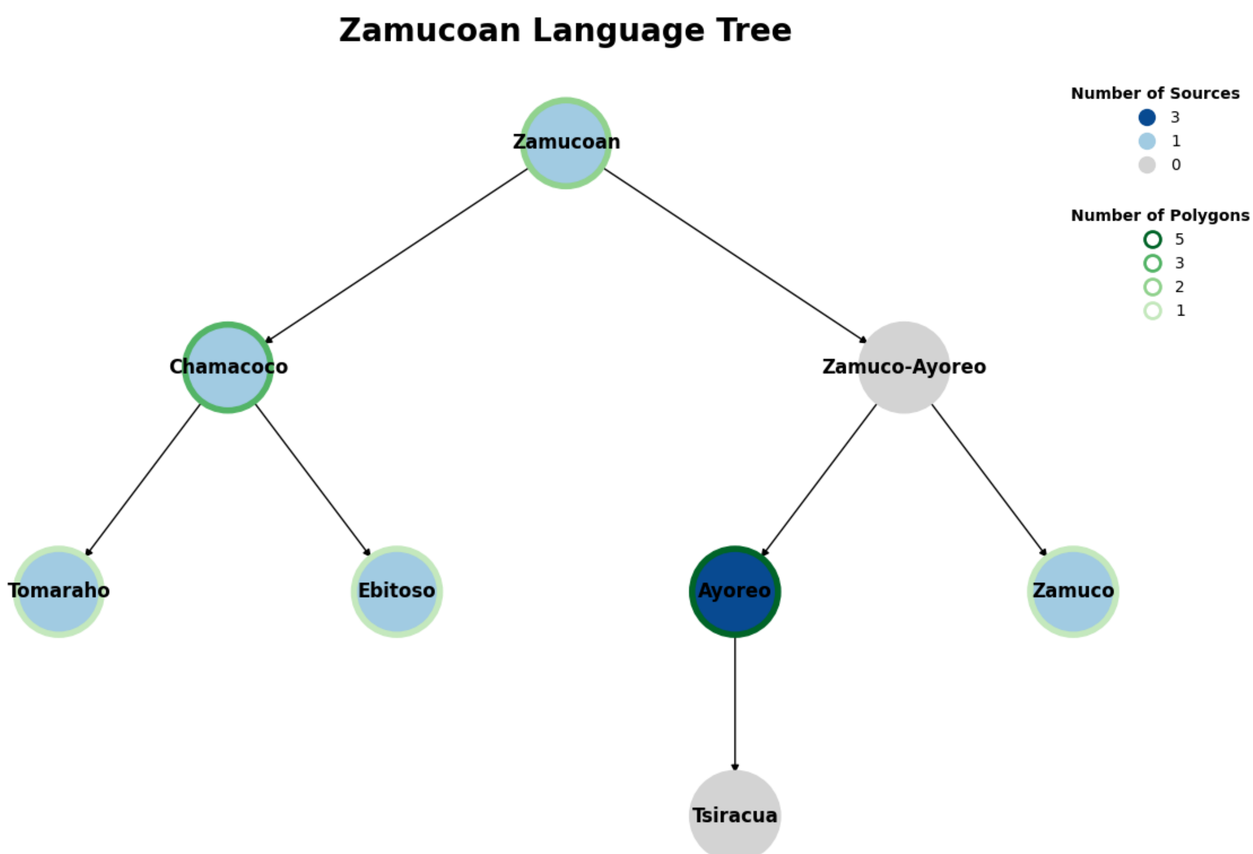


Figure 2 *- Zamucoan Language Tree with Data Availability for each Languoid.*

As the Figure 1 demonstrates, the overlap in language ranges across South America presents a significant challenge in terms of data representation and visualization. This complexity is further elaborated in the subsequent Figure 2, which focuses on the Zamucoan language family. Serving as an example of the fundamental structure of language distribution, this language family effectively showcases nearly all the challenges associated with depicting language diversity. The visualization of the Zamucoan family highlights the need for adaptable and detailed strategies in mapping languages, addressing issues such as overlapping territories, distinct dialects, and the impact of language extinction on available data.

This tree diagram exemplifies a typical mid to small-sized language family, represented by interconnected nodes that each signify a specific languoid. The lines connecting these nodes illustrate the linguistic lineage, while the size of the nodes is proportional to the number of spatial polygons and sources documenting each languoid. The representation highlights several challenges inherent in visualizing linguistic data. For instance, some languoids lack spatial data entirely, as indicated by smaller nodes without an accompanying polygon. This absence of spatial representation can be due to various reasons, such as a scarcity of research or the extinction of a language without sufficient geographical documentation.

On the other hand, there are nodes containing multiple polygons, often resulting from the same languoid being documented in several source maps. This multiplicity of representations within a single language family emphasizes the necessity for a nuanced approach in platform design. It is crucial to ensure that each languoid's lineage and spatial presence are accurately portrayed, despite variations in data volume and source diversity. Such careful consideration is essential when designing a platform to effectively display these linguistic relationships and territories, particularly when not every ancestor or descendant languoid has a clear or consistent spatial representation.

The diagram thus not only provides insights into the Zamucoan family's structure but also reflects the broader challenge of designing data visualization systems that can accommodate varying levels of detail and completeness in linguistic data. It emphasizes the importance of flexibility and adaptability in the platform's architecture to capture the full spectrum of language diversity and lineage, from well-documented languages with extensive spatial data to those with limited or no such information.

3.4.3 Spatial Distribution of Language Family

Following the exploration of the Zamucoan language family tree, the accompanying Figure 3 provides the spatial representation of the same family. This map displays the polygons that correspond to the languages within the Zamucoan family, as identified by the tree structure previously discussed.

The Figure 3 reveals again substantial overlap between polygons, illustrating the complexity of geographical boundaries among these languages. It's evident that the spatial extent of each languoid does not necessarily align with its position or level in the hierarchical structure of the family tree. Additionally, it is not a consistent pattern for a parent languoid to spatially encompass the ranges of its descendant languoids.

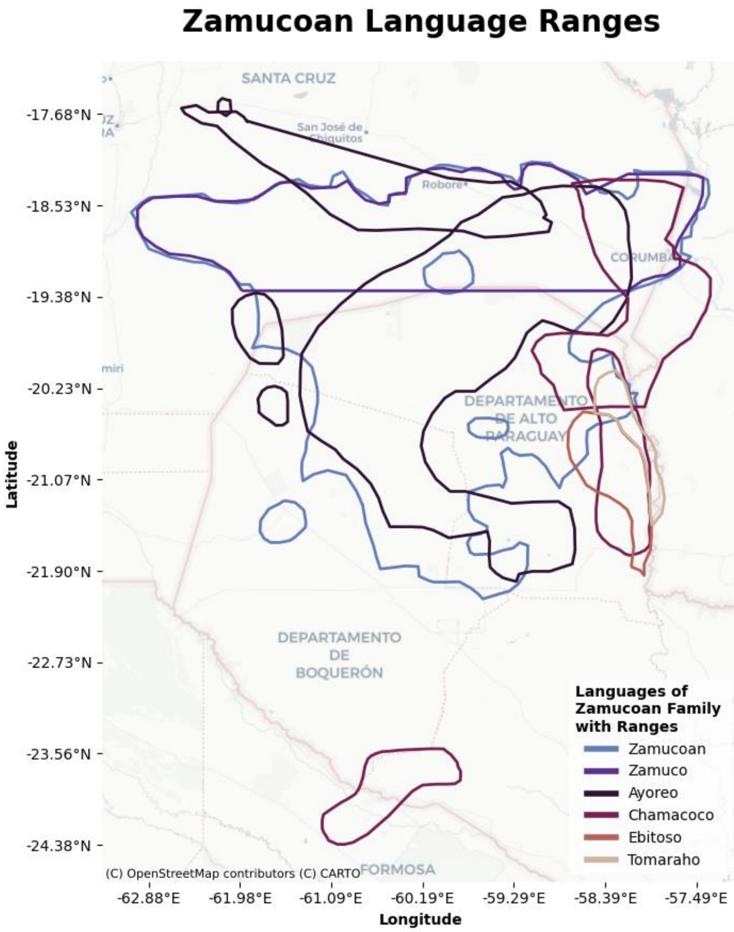


Figure 3 *- Geographical Extent of Language Ranges of the available Zamucoan Family Data.*

Such observations suggest that the geographical spread of languages within a family can be complex and is influenced by various factors that may not be immediately apparent from genealogical data alone. The task of accurately displaying these overlaps and divergences presents a challenge for the design of the visual platform, requiring an approach that can adequately represent both the linguistic relationships and the spatial data.

3.5 Challenges from the Data

In the data section of this study, it was shown the complex nature of the linguistic landscape in the Glottography database through the examination of South America and its associated spatial data. Several challenges have been identified through the analysis of the polygon and map datasets that present opportunities for methodological refinement.

Firstly, the analysis has revealed significant overlap within the language range polygons, which underscores the intricate nature of linguistic territories and the need for adequate data representation strategies. This complexity is magnified by the discovery that, in some cases, multiple languages share geographical spaces, creating a dense linguistic diversity.

Secondly, the representation of the Zamucoan language family has highlighted the disconnect between spatial and genealogical data. The spatial data does not always align with the hierarchical structure of language families, with the size of polygons not corresponding to the node level in the family tree and parent languoids not necessarily encompassing their descendant languoids.

Additionally, the presence of multiple polygons for single languoids, derived from various source maps, introduces a level of complexity in data management and necessitates a robust platform capable of reconciling these variations.

Finally, the scalability of the visualization platform is a critical challenge. The current dataset, while substantial, is only a representation of a much larger global linguistic dataset. The platform must, therefore, be designed to accommodate an expansion of data, ensuring that the methodologies applied can handle a broader scope without compromising on the accuracy and detail required for linguistic research.

4. Technical Setup

The development of the Glottography, initiated two years before the start of this thesis, is foundational to understanding the contribution of the current work. Early development benefited significantly from Gereon Kaiping and Peter Ranacher's efforts in backend setup and database schema development. The initial selection of a SQLite database facilitated simple data management and early prototyping. However, to meet the project's growing needs for geospatial capabilities, a transition to a PostgreSQL database equipped with the PostGIS extension was made. This shift was crucial for enhancing the storage, indexing, and querying of geospatial data, aligning with the project's objectives. Additionally, the adoption of DBeaver as a database management tool played a key role, providing a platform that supported an intuitive engagement with the project's data structures.

Originally, the project adhered to a monolithic architecture within the Pyramid framework, encompassing both frontend and backend functionalities. This setup, while functional, soon revealed limitations in scalability and flexibility, requiring a strategic shift. Familiarity with the project before the thesis had revealed its intricacies and potential, leading to advocacy for adopting a client/server model. This new architecture was aimed at enhancing modularity, maintainability, and the capacity to utilize specialized technologies for different aspects of the project. With the backend firmly rooted in the Pyramid framework, renowned for its flexibility and robustness, a standard process for comprehensive data handling was engineered. This process involves extracting data from the database, leveraging SQLAlchemy as the Object-Relational Mapping (ORM) tool to model this data for further processing. SQLAlchemy plays an important role in abstracting the database operations, allowing for seamless interactions with the database using Python objects instead of raw SQL queries. This ORM significantly enhances the development efficiency and maintainability of the database interactions, ensuring that the data is accurately modeled and transformed into structures conducive for frontend consumption.

4.1 Frontend Development

Taking the lead on frontend development, the choice fell on the Next.js framework, motivated by prior experience with React and the extensive support from its community. This decision marked a significant phase in the project, focusing on creating a dynamic and interactive user interface. A key component of the frontend architecture is the integration of the deck.gl library. Deck.gl emerges as an optimal choice for the Glottography, primarily due to its adeptness at handling large volume polygonal datasets. Its core strength lies in high-performance rendering, leveraging WebGL to tap into GPU capabilities for fast and efficient visualization of complex geometries. This is particularly crucial for rendering intricate polygonal shapes inherent in linguistic and geographical data, ensuring smooth user interactions even with extensive datasets.

Designed with geospatial data in mind, deck.gl offers specialized layers, like the GeoJsonLayer, making it well-suited for accurately and attractively visualizing spatial representations of language ranges. Its scalable architecture supports dynamic data management techniques, facilitating the growth of the Glottography's dataset without sacrificing performance.

Deck.gl's support for interactive and dynamic visualizations enhances user engagement, allowing for real-time updates and interactions such as zooming and panning without any movement stutter. This synergy not only enhances the utility and user experience of the web atlas but also underscores deck.gl's suitability for projects demanding geospatial data visualization.

4.2 Architecture:

4.2.1 Backend Group

Database (SQLite/PostgreSQL + PostGIS): This is the starting point of the backend, where all the geospatial data relevant to the project is stored. Using PostGIS with PostgreSQL (or SQLite for simpler setups) is a common practice for handling complex geospatial queries efficiently, making it a sensible choice for the foundational data layer.

Pyramid (Backend) + SQLAlchemy: Pyramid, a Python web framework, is suitable for serving dynamic content and handling web requests. SQLAlchemy, an Object-Relational Mapping (ORM) tool, facilitates database interactions by translating Python classes to database tables and automatically converting function calls to SQL queries. This setup efficiently bridges the gap between the complex data stored in thedatabase and the web server logic implemented in Pyramid, illustrating a standard practice in web application development for data-driven projects.

4.2.2 Frontend Group

Next.js (Frontend) Server: Next.js is chosen for its advantages in server-side rendering, static site generation, and overall React framework compatibility, offering a modern approach to building interactive user interfaces. The server generated by Next.js serves the processed content to clients, facilitating fast load times.

deck.gl + Other Libraries: For visualizing complex geospatial data on the web, deck.gl provides high-performance rendering capabilities, making it an ideal library for the project. Its use, alongside other frontend libraries, enhances the application's interactivity and visual appeal, catering specifically to the demands of rendering large geospatial datasets smoothly.

Browser (Web Application): The end point of the user interaction, where the web application is accessed and rendered. This component represents the user-facing side of the application, making it a critical piece of the architecture for evaluating the project's success in delivering a seamless and informative user experience.

The depiction between backend and frontend groups, connected through an API layer, is an approach in web application design, emphasizing separation of concerns, modularity, and scalability. This structure not only facilitates development and maintenance but also supports efficient data handling and user interaction, aligning well with the project’s goals of providing an interactive and informative web atlas. In summary, the content and the way it's structured in the visualization accurately reflect the logical setup for a web-based application like the Glottography, showcasing an architecture that leverages modern technologies and practices for efficient, scalable, and user-friendly application development.

5. Methodology

Shifting from the technical underpinnings to the methodological process of Glottography, it is important to outline the UCD approach. Presenting this structure early helps maintain clarity throughout the methodology section. While each phase of the UCD process is significant, this thesis emphasizes particularly on interaction and usability studies, along with conceptual development and prototyping. These phases are discussed in depth, highlighting how decision-making processes and user feedback drive the project forward. Less focus is given to the steps of implementation and debugging. This selective focus ensures a clear narrative that links each phase directly to its impact on user interaction and usability. By stating this structured plan at the beginning, the methodology section aims to present a logical and methodical narrative, reflecting the systematic progression towards a user-centered visual platform as seen in Figure 4.

Guided by the UCD framework, the workflow of the visual platform project for exploring spatial language data integrates user insights at every stage, from initial concept to prototype evaluation. Here's how the UCD process maps onto the project workflow:

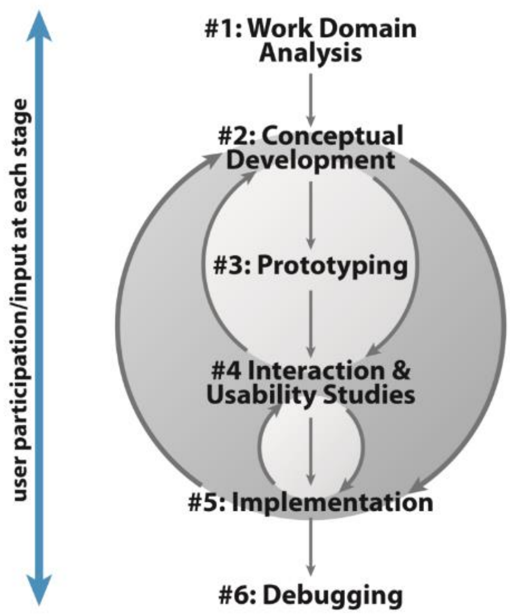


Figure 4 *- UCD Steps for Interactive Map Development (Robinson et al., 2005).*

5.1 Work Domain Analysis

The primary objective in this initial phase is to establish how users might interact with the platform and understand the linguistic data presented. The phase involves crafting a strategy to gain a comprehensive understanding of linguistic researchers' needs and preferences. To accomplish this primary objective, a two-step approach is adopted:

First, an initial prototype with basic interactive map capabilities, such as zooming, panning, and data filtering options, plus an information panel is developed. This prototype aims to present language ranges on the map and provide additional linguistic information when interacted with. It serves as a preliminary version to gauge potential functionalities and collect user feedback.

Second, an interview guideline is developed to gather user requirements and needs from qualitative interviews with potential users (see Appendix A). This guideline is aimed at collecting detailed feedback on the prototype's usability, the displayed information's relevance, and the overall user needs. The goal is to engage the user in interactions with the platform and identify discrepancies between user expectations and the prototype's offerings.

5.1.1 Development of basic prototype for the qualitative Interview

The initial prototype of the Glottography drew insights from the URHIA project, where the categorization of platform functionalities into five distinct areas, platform type, layer management, spatial data query, cartographic map elements, and presentation of attributes, provided a clear direction for development (see Section 2.5.). This categorization crafting a functionally rich prototype that was also user-focused,A rource map addressing the specific needs identified in previous successful projects, like the URHIA or the Native Digital Land.

The development of the Glottography prototype targeted technical and user interface challenges, such as the handling of overlapping polygons and the integration of diverse data sources, key issues also addressed in the creation of URHIA. These foundational tasks were critical for presenting clear and accurate linguistic data. Incorporating key navigational features, the prototype enhanced map navigation and filtering capabilities, allowing users to uncover data with greater efficiency. Visual design choices, informed by established cartographic best practices, utilized color coding and highlighting to differentiate and clarify linguistic data visually.

The development process followed a pragmatic, ad-hoc "Quick and Dirty" approach (Sharp et al., 2007), guided by existing knowledge and immediate expertise, mirroring the iterative strategies leveraged in successful spatial data platforms. This approach placed a premium on rapid prototyping and iterative adjustments, relying heavily on direct user feedback to refine and adapt the prototype. Such a strategy ensured that development remained flexible and responsive, allowing the prototype to evolve in alignment with user experiences and feedback, thus setting a strong foundation for the subsequent qualitative interviews that would further shape the Glottography platform.

5.1.2 Addressing the Challenges from the Data Chapter

Spatial overlap of language polygons pose a foundational challenge (see Section 3.3.). It is not possible to showcase all language polygons at once since the display window would be cluttered with overlapping polygons. Heavy overlap causes difficulties to select certain polygons and it leaves almost no room for labeling or for color coding to be effective for the user to recognize what languoid the polygon actually represents.

To address the challenge of reducing polygon display while preserving data integrity, two preliminary methods were considered, each with inherent drawbacks. The first method involves aggregating and generalizing polygons based on shared glottocodes, simplifying geometry for a broader overview, and integrating name labels for identification. However, this approach compromises data accuracy by merging distinct sources without accounting for their varied mapping methodologies. Such aggregation can significantly distort the representation of language areas, as it overlooks the definitions of linguistic territories, leading to potential misinterpretations of spatial relationships.

The integration of expert-based decisioning presents an approach that leverages specialized knowledge to refine data aggregation processes, as was seen in Rantanen *et al*. (2022). By involving experts in language families or regions, this method aims to enhance the accuracy of generalized data, ensuring that linguistic distinctions are adequately preserved and represented. Experts could critically evaluate how language data is collected and aggregated, making informed decisions that prevent the oversimplification and potential distortion of language territories. While this expert-driven strategy offers a promising solution to maintain data integrity, it requires access to a specialized knowledge that is currently beyond the scope of this thesis. Therefore, while recognizing the value of expert involvement, the current project will focus on developing and optimizing alternative methodologies that are more feasible within the existing capabilities, yet still strive to address the complexities inherent in accurate language data representation.

An alternative method, tessellation, applies geometrical segmentation to consolidate polygon counts on a global layer. This technique aids in analyzing intersecting polygon areas, offering insights into the aggregation of language polygons within each tessellated segment, similar to the modifiable areal unit problem (Wong, 2004). However, tessellation inherently overlays geometric segments on existing data, which can obscure crucial details and simplify unique linguistic characteristics. Tessellation tends to simplify semantic depth by aggregating distinct linguistic features into broader categories, which can dilute the precision of the data . Although tessellation helps manage large datasets by reducing visual complexity, its tendency to generalize critical details makes it suboptimal for projects requiring accurate depiction of linguistic diversity. However, as techniques improve and new methodologies are developed, tessellation may become a viable option in the future, especially if advancements enable better preservation of underlying data characteristics. Therefore, it remains a consideration for future projects where enhanced methods might overcome current limitations.

Given these considerations, the most effective approach for the initial prototype prioritizes the direct display of individual languages or specific source maps. This strategy involves showing only one polygon or source map at a time, effectively highlighting each language entity independently. A source map, that comprises of all the polygons originating from a particular source, allowing for a clear and focused engagement with each dataset without the complications associated with data aggregation and overlap. By adopting this method, the prototype enables a straightforward solution for visualizing linguistic distributions. Users can engage with a singular, unambiguous dataset at a time, which circumvents the issues of overlapping and merging data from multiple sources. For conducting comparative analyses, users can open separate browser instances to view different datasets concurrently, thereby maintaining the accuracy and integrity of each map's representation. This sequential display method remains beneficial for qualitative examination, offering the flexibility to toggle among various sources as needed. It enhances user interaction by providing a clear method for detailed scrutiny of individual datasets, avoiding the distortions typically introduced by more complex data aggregation techniques.

Incorporating the hierarchical structure of language families is essential for the prototype. Mapping a complete family tree faces spatial overlap challenges. Initially, these were overlooked to prioritize a holistic family view, deemed valuable for eliciting user insights. To address this, a non-spatial representation of language family hierarchies within the interface is proposed. This involves integrating a hierarchical tree model, facilitating an understanding of languoid relationships and enhancing navigation within language families. This model, similar to computer file system structures, offers a familiar and efficient means for visualizing and exploring linguistic dependencies.

5.1.3 Layout

In conceptualizing the initial interface prototype, a pivotal strategy included partitioning the user interface into two principal components: the information panel for non-spatial data and the map for spatial data visualization. This division is instrumental in facilitating user navigation and establishing an organized structure, enabling users to engage more effectively with each component. For example, users can manipulate the map through actions like zooming and panning, all while maintaining uninterrupted access to the linguistic data displayed in the information panel. This division of information display, a common technique in well-established applications like Google Maps[5], OpenStreetMap[6], and MapQuest[7], was informed by the design guidelines gleaned from an evaluative study of Web mapping sites (Nivala et al., 2008).

These guidelines underscore the importance of an intuitive layout where the user should be able to start using the map immediately upon entering the page. This prototype's design reflects this by adopting a top-to-bottom, left-to-right flow in the interface layout, resonating with Western reading habits and established navigation schemas (Granlund et al., 2001). This familiarizes users with a seamless and accessible entry point to the platform's information and functionalities. The layout's alignment with these practices is visually represented in Figure 5, showcasing how the information and map components are organized along the left-to-right axis, with details within the information panel arranged from top to bottom.

The information panel is segmented to separately display language details, sources, and lineage information, which aligns with the guideline to group UI elements logically. This organization allows users to efficiently toggle between views tailored to their specific information needs. The map component's primary focus is the clear depiction of the spatial distribution of languages, a crucial aspect supported by the guideline advocating for a search box with a principal role in the layout to enhance user experience. The synergy between the information and map components is maintained through reactive adjustments of content based on user interactions, such as selecting a language polygon on the map or switching between the language and source view in the information component.

By integrating these design principles, it is ensured that the prototype not only presents linguistic data accurately but also offers a user-friendly interface. This attention to a structured and intuitive design allows for a more engaging and efficient user experience, reflecting a deep consideration of usability aspects highlighted by the recent study.

[5] https://www.google.com/maps/
[6] https://www.openstreetmap.org/
[7] https://www.mapquest.com/

The interface's **language section** is systematically structured to display essential linguistic data. It includes the primary name of the language, any known alternative names, and a direct link to the corresponding Glottolog page, referenced via the language's glottocode. This section also includes represents the hierarchical tree model, outlining its phylogenetic placement within related languoids, and lists different source maps where available. As it addresses the previously noted challenge of polygon overlap, this section displays a single language's spatial range to ensure clarity and precision in the representation on the map component.

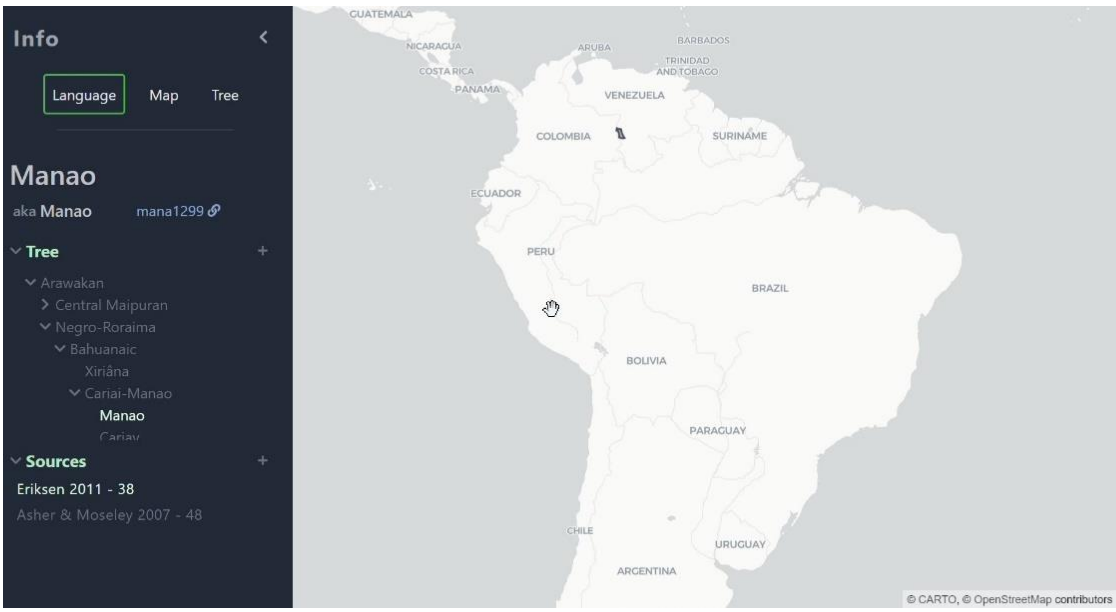


Figure 5 *- Language View in the Glottography Application with the Language Section on the Left Side.*

The **map section** of the interface is organized to provide an overview of the linguistic data derived from specific sources (see Figure 6). At the top of the source map panel, the author and publication year of the source are displayed, followed by a unique identifier for the dataset. It also details the figures or pages within the source where the data can be found, and indicates the availability of an image or visual representation of the source material. In the map component, polygons are displayed, each corresponding to the language data referenced in the information panel. This design adopts a 'one at a time' approach, showcasing data from a single source at any given moment, which minimizes the issue of overlapping polygons and aligns with the strategy to enhance data clarity in the visualization.

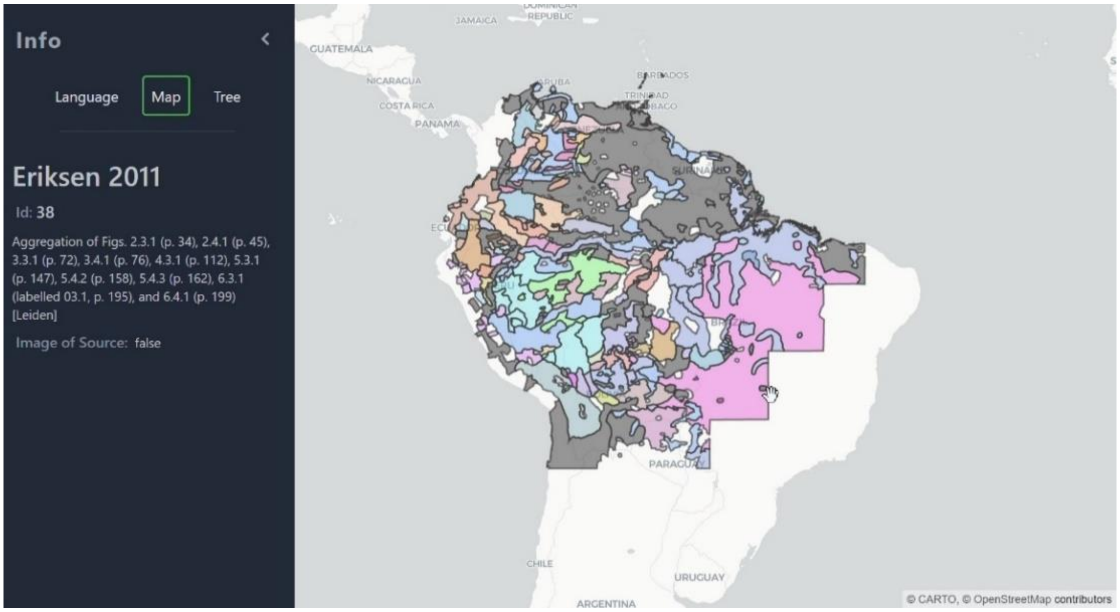


Figure 7 - *Map View displaying all Polygons from the Eriksen Source Map in the Glottography Initial Prototype with the Map Section on the Left.*

In the **tree view** panel of the interface, the focus is on representing the genealogical structure of language families, mirroring the tree representation already introduced in the language panel (see Figure 7). This view provides an expandable tree that delineates the lineage relationships among languages, offering a hierarchical perspective. Correspondingly, the map displays a collective set of polygons representing the geographic distribution of all languages within the selected lineage being the currently selected language up to the most common ancestor. This design choice facilitates a visual understanding of the relationships and spatial distribution among related languages, serving as a valuable tool during user interviews to elicit feedback on user needs and desires regarding the visualization of language family.

Figure 6 - *Tree View showing the Lineage from the Selected Language in the Glottography Initial Prototype.*

5.1.4 Navigation

Click Behavior: The implementation of a point-in-polygon algorithm is central to user navigation, triggering feedback via pop-ups on the map component when a user clicks within a polygon. This pop-up, see Figure 8, displays a list of polygons detected by the algorithm, consolidates the selection process. When a single feature is identified, the system foregoes the pop-up, directly highlighting the polygon and updating the language information in the information component. This approach ensures that user interactions are both responsive and informative, enhancing the usability of the map interface.

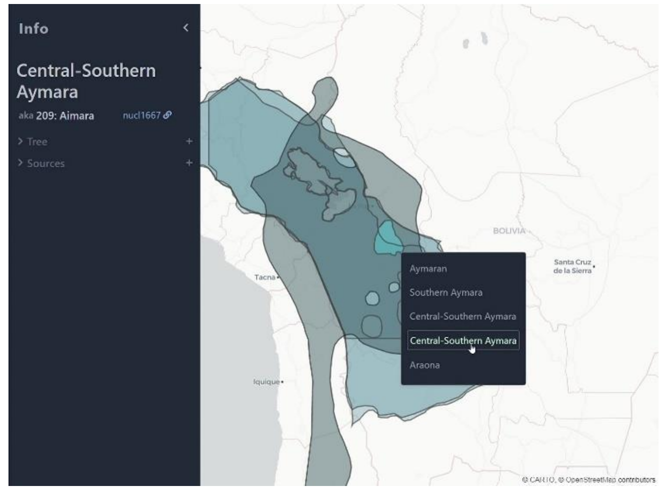


Figure 8 *- Snapshot of the Initial Prototype with the Point-In-Polygon Selection.*

Language Tree for Navigation: The language tree serves, see Figure 9, as a navigation tool, where each node's selection updates the displayed language. Clicking on a languoid within the language tree changes not only the language details in the information component but also, when applicable, switches the corresponding polygon on the map to reflect the selected language. This feature establishes a direct link between the hierarchical structure of language relationships and their geographical representation, enabling users to explore linguistic connections both spatially and genealogically.

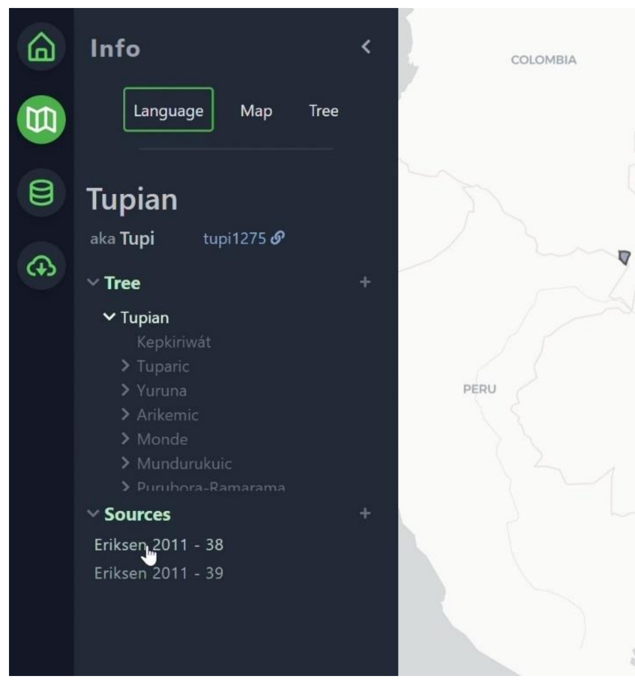


Figure 9 - *Snapshot of the Initial Prototype with the Showing the Possible Navigation Functionality in the Language Tree and the Source Dropdown Menu.*

Source Selection: Incorporating a source dropdown menu, see Figure 9, within the language information component allows users to refine their exploration based on the data source. Selecting a different source than the currently highlighted one triggers a switch in the displayed polygon on the map, corresponding to the newly selected source. This functionality supports detailed source-specific investigations, offering users the flexibility to compare language data derived from different research or documentation.

Switching Between Views: The ability to switch between the language view, source view, and tree view is integral to the platform's navigational framework, see Figure 10. Transitioning from the language view to the source view prompts the system to identify the polygon currently displayed on the map, determine its originating source map, and then display all polygons associated with that source. Conversely, switching to the tree view presents the lineage of the selected language, extending up to the most common ancestor within the language family if available. This versatile navigation system enables users to seamlessly transition between different perspectives of linguistic data, facilitating a comprehensive exploration of language diversity and distribution.

Figure 100 *- Snapshot of the Initial Prototype displaying where the Views within the Application can be Changed From Language to Map to Tree View.*

Datatable: The platform incorporates a datatable view accessible via a distinct page within the Glottography website, separate from the main map interface, see Figure 11. This dedicated page is populated with essential non-spatial language data, including language names, families, glottocodes, and sources. Users can navigate through this data to refine their search and exploration. The separation of the datatable from the map interface means that users need to switch to the datatable page whenever they wish to browse through non-spatial data. Upon selecting an entry from the datatable, the system automatically transitions the user to the map interface, where the specific language information corresponding to the selected entry is visually highlighted. This feature not only enhances user interaction by integrating detailed textual data with spatial visualization but also encourages a comprehensive approach to exploring language diversity and distribution, albeit with the necessity of toggling between views.

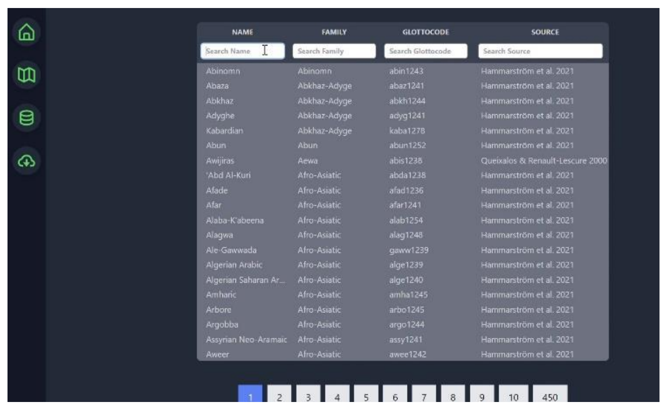


Figure 11 *- Snapshot of the Datatable in the Data Page within the Glottography Initial Prototype.*

5.1.5 Map feature representation

Base Map: The Carto's Positron base map, chosen from a limited selection available through the DeckGL library, served as the foundational layer. This base map was selected for its clean and unobtrusive design, which supports the overlay of additional data without overwhelming the user visually.

Polygon Design: Polygons, representing language areas, were designed with thin black borders and no labels to maintain a clean map interface. A hover effect was implemented to display the name of the language, enhancing user interaction without cluttering the map. The fill color of each polygon was carefully considered to ensure that languages in close proximity had distinguishably different colors for easier identification. However, achieving distinct color coding for all polygons proved challenging due to the limitations in color variance, occasionally resulting in less distinct color differentiation. Transparency is applied to the polygons to facilitate the observation of overlapping language distributions, ensuring that the spatial intersections of linguistic territories are perceivable while maintaining the visibility of the underlying cartographic details.

Highlighting Selected Polygons: When a polygon was selected, it was highlighted using a different border color and an overlay that subtly darkened the color hue of the selected language area. This method made it relatively clear which language was currently selected, though it was acknowledged as not the optimal solution. Despite its limitations, this approach was deemed fully functional for the initial prototype iteration, allowing for clear visualization of selected languages while recognizing the need for further refinement in future versions.

5.2 Interview

Upon finalizing the prototype, featuring core interactive map functions, the subsequent step involves creating an interview guideline. This guideline is structured to elicit feedback on the prototype, grounding the discussion in user interaction and experience with the initial design. This transition emphasizes a methodical approach to refining the platform based on specific user feedback.

5.2.1 User interviews as a method for requirement elicitation

Requirements elicitation is key aspect to reveal requirements insights from key stakeholders. In this project the key stakeholders will be mostly the end-users, and therefore it is vital to contact comparative and general linguists and inquire on their needs and requirements for the data visualization in the map interface (Goguen and Linde, 1993). Interviewing the stakeholders is a common technique to gather the requirements from users and demands the interviewer to be a facilitator, listener, and guide throughout the interview . Intricate knowledge on the project's field of study as well as the interviewees and their respective academic interests allows to get an in-depth perspective from the relevant stakeholders. Respecting guidelines for qualitative interviews is key to conduct an successful interview and that's why relying on strategies such as given by the Harvard Faculty of Sociology[8] can be helpful.

This interview should ultimately support the documentation of the user requirements, which will create an overview of the general needs and scope of the project itself. The results can the later be used to create user stories which are a tool to describe a desired functionality of a system from the user's point of view. User stories allow for user-centered content creation, strategy refinement and a collaborative content development for the developer (Lucassen et al., 2016). Furthermore, they help to simplify the core requirements and break them down to an easy-to-follow approach to support the developer when realizing the software prototypes/end-product.

Knowing how integral the requirements are for not only the first steps of the development process but also the subsequent ones, it is of critical interest to thoroughly create a guideline for the interview to gather the most relevant user needs and requirements. Including the above mentioned strategies the results for the guideline are divided up into 3 main categories, namely background, data exploration/ map interface, personal research/ data visualization. Each category has main questions which are expanded with follow up questions to additionally add research depth if required.

[8] https://sociology.fas.harvard.edu/files/sociology/files/interview_strategies.pdf

The guideline structured the interviews around three main categories, Background, Data Exploration/Map Interface, and Personal Research/Data Visualization, each chosen to cover a broad spectrum of user interactions and needs. These categories were further divided into subtopics, ensuring a comprehensive exploration of user experiences and preferences:

1. **Background:** Questions aimed at understanding the user's professional background and prior experience with language visualization tools. This insight helps align the interface with user expectations and existing knowledge, making the tool more intuitive and accessible.
2. **Data Exploration/Map Interface:** This section focused on initial impressions of the prototype and specific functionalities users deemed essential for their work. Feedback gathered here is critical for identifying which features require prioritization and further refinement to enhance the user experience.
3. **Personal Research/Data Visualization:** By understanding how users intend to utilize the map interface for their research and what visualization tools they value, the development can focus on incorporating features that support these activities, such as the integration of hierarchical linguistic data in a manner that is both informative and easy to navigate.

The rationale behind each category and question within the guideline was to ensure that every aspect of the prototype's development, from its initial design to subsequent iterations, was informed by an understanding of the end-user's needs.

The table outlines a systematic approach for gathering user requirements through interviews, crucial for refining the prototype's design. It divides the interview framework into two main types: "General" for overarching insights and "Application Specific" for focused feedback, concluding with a "Finalizing" stage for additional reflections.

In the "General" segment, interviews probe users' backgrounds, ensuring the prototype matches their experience and expectations. For example, a linguist with extensive fieldwork might require different data visualization features compared to a theoretical linguist. Similarly, understanding users' prior exposure to language visualization tools allows for tailoring the prototype's complexity. A user familiar with advanced mapping software might seek more sophisticated features than one used to basic tools.

The "Application Specific" segment collects initial user impressions to evaluate the interface's usability, pinpointing areas for immediate enhancement. It also identifies the particular language information users value, such as dialect distribution or language endangerment levels, to prioritize relevant features. Preferences for visual representations, like color coding for language families or interactive maps showing language spread, guide the interface's design. The inclusion of hierarchical linguistic data is informed by users' interest in exploring language genealogies. Additionally, assessing the importance of scholarly references shapes decisions on integrating academic resources directly into the map.

Finally, the "Finalizing" segment emphasizes the significance of capturing overarching feedback and any unresolved user needs. This continuous feedback loop is essential for iterative development, ensuring the prototype evolves in response to real user interactions.

Table 4 – User Interview Structure with Categories and Rationale for Clarification of Inquiry

Type	Category	Rationale
General	Background	Understanding users' backgrounds and prior ensuring the interface is aligned with user expectations and prior knowledge.
	Language Visualization	Gathering insights into users' familiarity with language visualization tools helps in adjusting the prototype's complexity and functionality to match user expertise.
Application specific	Initial Impressions	Feedback on the prototype's initial design provides a baseline assessment of the user interface, which is critical for iterative design improvements.
	Desired Language Information	Identifying which details users find essential allows for a data-driven approach to feature prioritization, ensuring that the most valued information is accessible.
	Views and Perspectives	Exploring user preferences for visual representations guides the aesthetic and functional design of the map interface, promoting a more engaging user experience.
	Language Family Tree	Understanding the importance users place on language relationships assists in deciding how to integrate hierarchical linguistic data in a user-friendly manner.
	References and Sources	Determining the value of references informs decisions on incorporating scholarly resources, enhancing the prototype's educational utility.
	User Interaction	Insights into desired interaction features shape the development of user interface components, fostering a more intuitive tool.
Finalizing	Closing and Feedback	Capturing additional user feedback ensures that any unaddressed needs or potential improvements are considered for future prototype iterations.

In the prototype's development, consulting a diverse range of stakeholders, specializing in historical linguistics, dialectology, sociolinguistics, and geospatial interdisciplinary research, was essential for a comprehensive requirement elicitation (see Appendix A). The inclusion of such a wide array of expertise not only enriched the development process but also ensured that the prototype would meet varied user needs effectively. Stakeholders from these distinct fields brought unique perspectives that were critical in shaping the prototype's functionality and usability.

The synthesis of these expert insights was instrumental in refining the prototype, making it a multifunctional tool tailored for linguistic research. By integrating a broad spectrum of functional requirements and preferences, the development process benefited from a depth of understanding that significantly influenced the prototype's design. This collaborative approach not only enhanced its applicability across academic and educational settings in linguistics and related disciplines, but also ensured that the tool was robust enough to handle complex analyses and user interactions. This broadens its potential impact and utility.

5.3 Conceptual Development & Prototype

Following the work domain analysis, the next phase focuses on conceptual development and prototyping, where the emphasis is on translating user requirements into tangible features for the application. This phase consolidates the insights from the requirement elicitation process, examining each requirement's feasibility, complexity, and necessity. By grouping these requirements, a clearer overview emerges, guiding the development of application features that directly respond to identified needs. This approach ensures a structured transition from abstract user needs to specific functionalities within the application. The categorization of requirements established earlier plays a critical role in this phase, as it enables a direct linkage between the articulated needs of users and the corresponding features designed to meet these needs. This process is essential for developing an application that is both functional and user-centric, ensuring that the final product aligns closely with the expectations and requirements of its intended audience.

5.3.1 Requirement Analysis

During the requirement analysis phase, the process begins with requirement elicitation, a stage dedicated to capturing and articulating user needs. Drawing from qualitative interviews and the detail within user profiles (see Appendix B), this stage distills the essential elements into preliminary requirements (Holtzblatt et al., 2004). User stories further refine this process by providing a user-centric narrative of desired system features (Lucassen et al., 2016). User profiles lay the groundwork for understanding how different users might interact with the platform, providing a rich context for the development of user stories (see Appendix B). User stories, expressed in natural language, specify the desired features from the standpoint of end-users.

When assessing the feasibility of these requirements, the focus is on their fit with the platform’s explorative purpose rather than just technical practicability (Berander and Andrews, 2005). For instance, a feature to analyze speaker counts isn't pursued at this stage, not because it's technically infeasible, but because the data is too sparse for meaningful exploration, as noted in Section 3.2.2.. This aligns with a prudent prioritization of features that can be confidently underpinned by reliable data. Similarly, functionalities that stray from the platform's primary explorative function, like the ability to download datasets, are designated as future improvements.

This shifts the development focus to requirements that not only align with the thesis’s goals but are also data-backed, marking them as actionable. These selected requirements are then prepared for the design and development phase, ensuring the platform remains dedicated to facilitating exploration. Requirements set aside for later stages are documented, paving the way for their possible future integration once the platform's explorative capabilities expand or as data availability improves. This process, including the decisions and steps leading to the prototype, is visually summarized in Figure 12, which illustrates the pathway from initial requirement elicitation through to the development of a functional prototype.

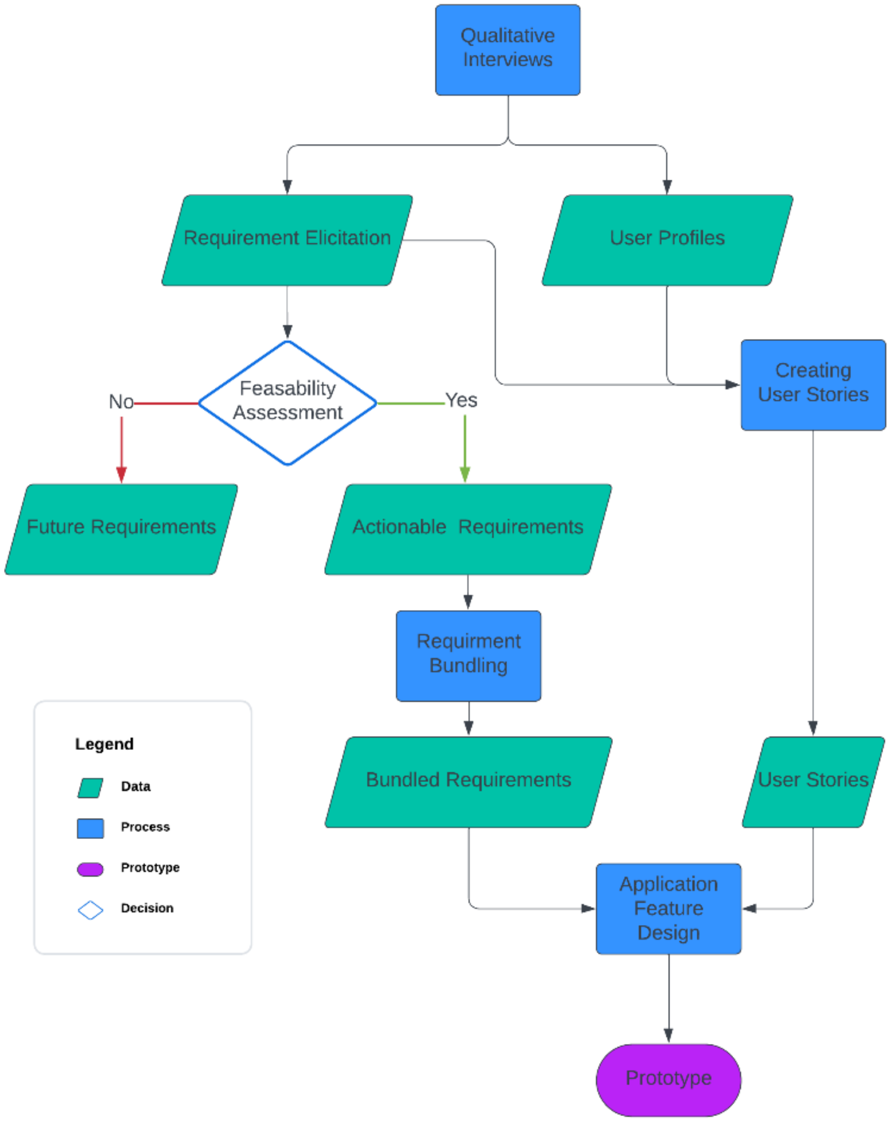


Figure 11 *- Schematic Representation of User-Centered Design Methodology, Depicting the Steps from Qualitative Interviews to Prototype Development.*

Grouping the actionable requirements based on shared functionality and user workflow enhances the development process by underscoring the collective importance of certain features. For instance, multiple requirements emphasize the need for interactive exploration tools, such as the ability to navigate through a language family tree. This convergence suggests a high value in developing this feature, further justifying its prioritization in design decisions. Furthermore, organizing requirements by their functionality and the user's workflow path. This not only supports the rationale behind design choices but also simplifies subsequent evaluations of the platform's usability. For example, assessing how effectively users can navigate through the language family tree to discover linguistic relationships becomes more straightforward when these actions are categorized under a unified workflow goal (see Figure 13).

Translating grouped requirements into application features is a critical step in the design process. It involves synthesizing requirements that share common functionality into cohesive features within the application. Taking the "Search and Real-Time Filtering" group as an example, the requirements to enable selection of languages through a fuzzy search and to filter data in real-time merge into the design of an intuitive search bar feature. This feature allows users to effectively filter through languages and obtain search results.

Figure 12 *- This Figure Illustrates the Breakdown of Grouped Requirements into Individual Requirements and their Corresponding User Interface Feature with the Example of Search and Real-Time Filtering.*

The articulation of grouped requirements into tangible features facilitates the creation of task-based scenarios (see Appendix C for all scenarios). These scenarios are instrumental for the upcoming user interviews, providing a structured way to assess the application's usability and the implementation of the features. They allow for an evaluation of how well the application supports user workflows, ensuring that the features derived from the grouped requirements indeed enhance the user's exploratory experience. The search bar exemplifies this transformation, where a set of related functional requirements is embodied in a single, user-centric application component, poised for evaluation in the subsequent interaction and usability testing phase.

5.4 Interaction & Usability Studies

This section of the thesis examines the interaction and usability studies conducted to evaluate the application. A diverse group of nine participants were engaged, primarily consisting of linguists with a keen interest in spatial data, along with a software engineer and a linguist specializing in digital environments. This deliberate mix aimed to improve how different user types engage with the application, drawing on the approach seen in the URHIA case study (see Section 2.5.) which utilized expert roles from various fields. Although simplified, this strategy preserved the inclusion of a varied group, each bringing a unique perspective to the study.

The participant pool for this interaction and usability study was diverse, ranging from domain professors to postdocs, PhDs, and master's students. Seven of the interviewees comprised linguists who are already familiar with geospatial language data. Their existing knowledge of linguistic connections and regional data sources is crucial, as it allows them to assess whether the application meets the established user requirements effectively. This familiarity enables them to focus more on evaluating the user interface and map interactions rather than spending time understanding the basic language domain concepts. In contrast, the inclusion of a software engineer and a linguist with web application expertise introduced perspectives from individuals less familiar with comparative linguistics. These participants primarily explored the application's functionalities, filling in domain knowledge gaps as needed. Their experiences are particularly valuable for identifying navigational intuitiveness and user interface efficiency without preconceived biases about the linguistic data.

Incorporating users with varying levels of familiarity with the linguistic content helps in understanding how different user groups navigate and interact with the application (Fox, 2015). It is evident from the initial interviews that experts in linguistic data sometimes focused excessively on the accuracy of the data representation, occasionally diverting attention from the core usability and functionality of the application. Conversely, those less familiar with the domain were more focused on the application's operational aspects, providing insights into user experience that were unaffected by the content's intricacies.

This mixed approach in participant selection ensures a comprehensive evaluation of the application, addressing both its functional efficacy and its content accuracy, while aligning the development closely with the needs and expectations of its diverse user base. The findings from these studies are instrumental in refining the application, enhancing its usability, and ensuring it serves the intended educational and research purposes effectively.

5.4.1 Task-Based Scenarios

Task-based scenarios typically offer a practical approach to conceptualize user needs indirectly, without directly consulting users (Sutcliffe et al., 1998). However, in the development of this prototype, these scenarios were utilized in a reverse manner to quantitatively assess whether the system fulfills user requirements (see Appendix C). By emphasizing the user's perspective through replicating real-life tasks, these scenarios ensure that the system's design aligns well with user goals. This alignment is crucial to UCD, as it confirms that user needs and behaviors are the primary influencers of design decisions.

Task-based scenarios provide a measurable framework to validate system usability, enabling the observation and analysis of user interactions in controlled (Carrol, 2003). By allowing users to execute specific tasks, designers gain the opportunity to identify obstacles and areas for improvement within the system. This approach fosters an iterative design process, which is a hallmark of UCD. The feedback generated from executing these scenarios is invaluable, as it directly informs subsequent design iterations, ensuring the development process remains focused on enhancing user experience. Ultimately, the application of task-based scenarios in this reverse manner underlines the principles of UCD by prioritizing a user-first approach, ensuring contextual relevance, and embracing the iterative nature of the design process.

To demonstrate how task-based scenarios are structured to validate system usability, an example will be shown here:

Searching and Real-Time Families

Task for Participant: "Suppose you're researching the 'Venezuelan-Cariban' language family. I'd like to understand how you would use the map features to focus on this family and analyze its distribution."

- **Checklist**
  - ✓ Find search and open it.
  - ✓ Participant enters a search term related to the 'Venezuelan-Cariban' language family in the search bar.
  - ✓ Participant selects the 'Venezuelan-Cariban' family from the search suggestions.
  - ✓ Review the responsiveness of the system to real-time filtering and the relevance of the filtered data.
- **Requirements Covered:**
  - Enable Selection of Languages via Fuzzy Search
  - Real-time Data Filtering (in search process)

***Figure 13** - Task Based Scenario. Searching and Filtering Language Families with the Task, the Checklist and the User Requirements that are Covered with that Scenario.*

This scenario operationalizes the application's search and filtering capabilities. Tasked with researching the 'Venezuelan-Cariban' language family, the participant employs the fuzzy search feature to locate this specific group within the platform's extensive linguistic database. Success is measured by the participant's ability to identify and select the language family amidst varied entries.

The user utilizes real-time data filters to refine the search, concentrating on the language family's geographic distribution. This process evaluates the application's capacity to dynamically represent search results, allowing immediate visual delineation on the map as parameters are adjusted.

This scenario, tailored to simulate a realistic research task, effectively tests two primary requirements: the selection of language families using fuzzy search algorithms and the application's ability to filter data in real time during the search process. It provides a focused context for assessing the integration and functionality of these key features within the user workflow.

5.4.2 System Usability Scale

The System Usability Scale (SUS), developed by John Brooke (1996), has emerged as an industry-standard tool in assessing usability. Its credibility and reliability in providing a 'quick and dirty' usability score make it an ideal choice for a variety of products, including the complex software tools used by linguistic researchers. The SUS is particularly advantageous because of its cost-effectiveness. As a free resource, it allows even the most budget-constrained projects to implement a robust usability assessment, which makes it perfect for this thesis.

One of the key strengths of the SUS is its efficiency. Consisting of only 10 questions, it respects the time constraints of participants, often yielding completion in just a few minutes. This brevity is crucial when dealing with experts in fields such as linguistic research, where time is a precious commodity. Despite its to the point nature, the SUS is renowned for its reliability. The SUS can effectively differentiate between usable and unusable systems, which is pivotal for iterative design processes (Bangor et al., 2009).

Given the scale's wide adoption, its scores can be benchmarked against established norms, providing a context for interpreting results. This benchmarking is critical in a field where precise and actionable usability metrics are valued. The adaptability of the SUS enables it to be applicable across diverse interface types, which is pertinent when considering the varied nature of linguistic research software (Sauro, 2011).

6. Results

This section outlines three core outcomes from the evaluation of the interactive prototype. Initially, the prototype's development and functionality are reviewed, establishing a foundation for further analysis. This is followed by the presentation of task-based scenario outcomes, which detail the capacity of participants to complete predetermined tasks, as captured by a structured checklist. The final portion of this section aggregates participant feedback to quantify the prototype's user experience through usability test results. Collectively, these findings yield an understanding of the prototype's performance in a controlled user study.

6.1 Final Prototype

The final prototype of the linguistic database, as presented in Figure 15, is the end result of a design process attentive to user feedback and project requirements. It encompasses several novel features resulting from a targeted user requirements elicitation process. The source code is stored on a Github repository[9] and is publicly accessible. In the Figure 15 a screenshot of the application in the language view mode can be seen.

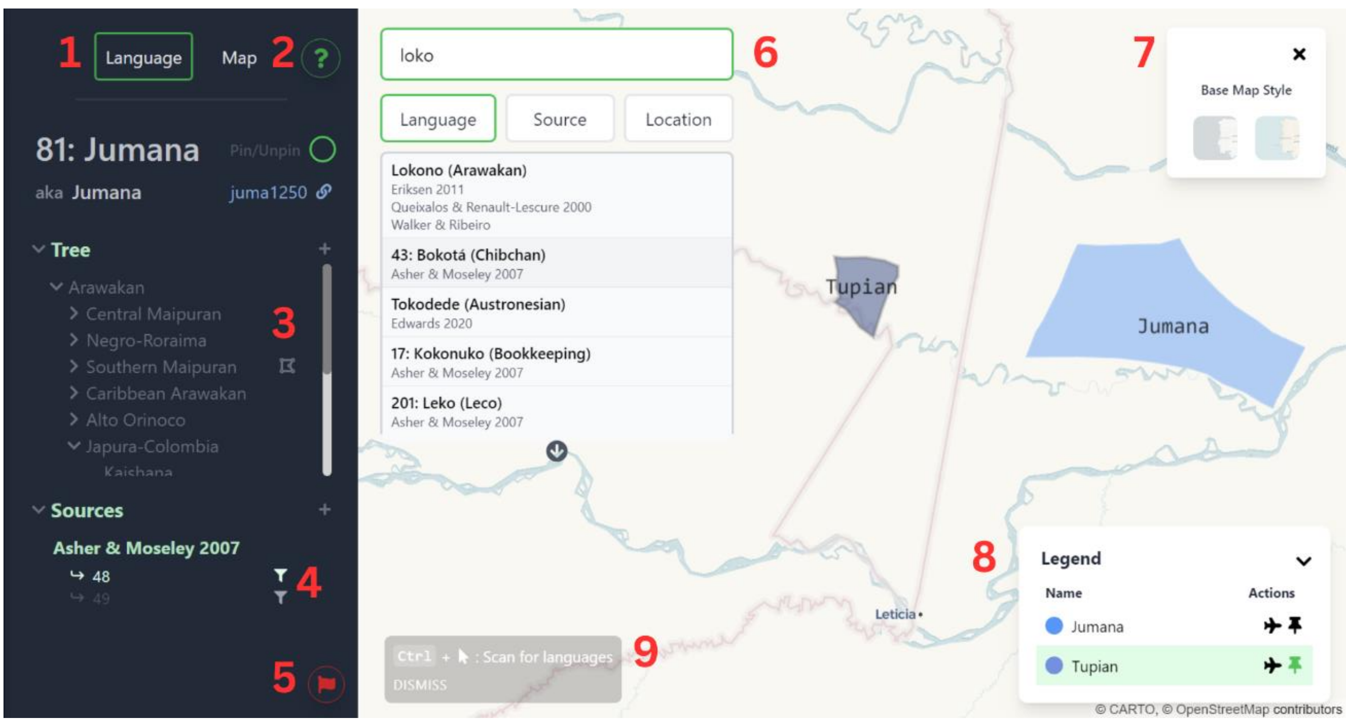


Figure 14 *- This Figure Presents the Final Glottography Prototype in the Language View, Highlighting New Features with Red Markers. These Markers Represent Functionalities for Visualizing Language Data Distribution.*

[9] https://github.com/Glottography/glottography_client

Table 5 *- Feature Table for Figure 15 with Feature Titles and corresponding Description*

Number	Feature	Description
1	View Selection Language/Map	Introduction of a toggle option to switch between language and map views. Removal of the obsolete tree view from the first interview round.
2	Navigation Guide	Added detailed descriptions of elements in the language view, including Glottolog links, Language Tree, and Source Options.
3	Language Tree with Indicators for Polygons	Updated logic for tree interaction: clicking the icon navigates to the languoid, while clicking the languoid name expands the tree.
4	Source Dropdown with Indicator	Included visual indicators to intuitively signal the ability to switch between different sources.
5	Report Form for Issues	Implemented a reporting feature for users to notify developers of misrepresentations, errors, or bugs.
6	Search Bar for Multiple Criteria	Replaced the old datatable with a search bar for languages, sources, and locations, and integrated an OSM API for location searches.
7	Base Map Style Settings	Added functionality for users to change the base map style, with plans for future relief map options.
8	Interactive Legend with Tools	Developed an interactive legend that allows users to pin and fly to language polygons on the map.
9	Map Click Behavior Indicator	Introduced a new click behavior for point-in-polygon operations to determine if a location falls within a language area.

Figure 15 showcases the language view within the visual platform, with a particular languoid highlighted for examination. The chosen languoid is distinguished by its name or alternative designation. Accompanying this is a Glottolog hyperlink, providing a gateway to the extensive linguistic reference database.

A phylogenetic display below the languoid information illustrates the taxonomic arrangement of the language family, allowing users to interactively explore the genealogical structure. An icon next to each languoid entry invites users to update the displayed data, ensuring they navigate within the language tree.

Below the phylogenetic representation, a source dropdown menu is positioned, enabling users to perform comparative studies by selecting among different scholarly interpretations of linguistic entities. This source menu is essential for examining the varying spatial hypotheses presented by distinct authors. Accompanying the source entries are small indicators to the right, highlighting the ability to switch between sources.

The interface features a search bar, streamlining the discovery of languages or sources. Given the potential unfamiliarity of users with the available sources, the source search function conveniently lists these resources, enhancing usability. Additionally, the search bar enables location queries, dynamically interacting with the map to reorient the viewport to the selected geographical point. This new search functionality replaces the previous datatable view on the other page (Section 5.1.4.). The integration of a search bar directly within the map interface significantly improves usability by eliminating the need to switch between pages to browse non-spatial data. Users can now access language and source information and perform searches without leaving the map view, providing a seamless and integrated user experience.

The current settings module is elementary, offering a singular functionality: the toggling between light and dark base map themes. Future iterations will expand this feature set to include a topographic relief option, enriching the visual differentiation of terrain.

The legend is a significant improvement, as it serves as both an index of visible polygons and an interactive tool. It provides at-a-glance verification of the polygons rendered on the map. Interaction with the legend entries via hovering prompts visual emphasis of the corresponding polygon, thereby aiding users in spatial identification. For more detailed examination, the 'FlyTo' feature recalibrates the viewport to center on the selected polygon. The capability to 'pin' polygons is particularly instrumental, as it retains selected polygons in the interface's memory, facilitating comparative analysis across disparate languages or sources without the constraint of lineage continuity. This feature proves invaluable when cross-referencing languages outside a shared phylogenetic branch.

It’s important to highlight two particularly noteworthy additions that, while not explicitly derived from initial user requests, are essential to the project's success. The inclusion of a report form for misrepresentations and bugs extends beyond specific user feedback. It is a strategic decision to ensure the quality and accuracy of the open-source project. This feature empowers users to participate directly in the project's improvement, fostering a collaborative environment.

Equally significant is the enhancement of the map's click behavior. Beyond the expected modifications, the prototype introduces an alternative way to switch between the interaction mode enabled by clicking on the map interface. By holding the 'Ctrl' key and clicking on the map, the user activates a scanning function that searches for language polygons in the vicinity of the clicked point on the map interface. This addition wasn't a direct user request, but rather an anticipatory feature designed to enrich the user experience by allowing for a quick switch between simple selection and a more detailed examination of language layers.

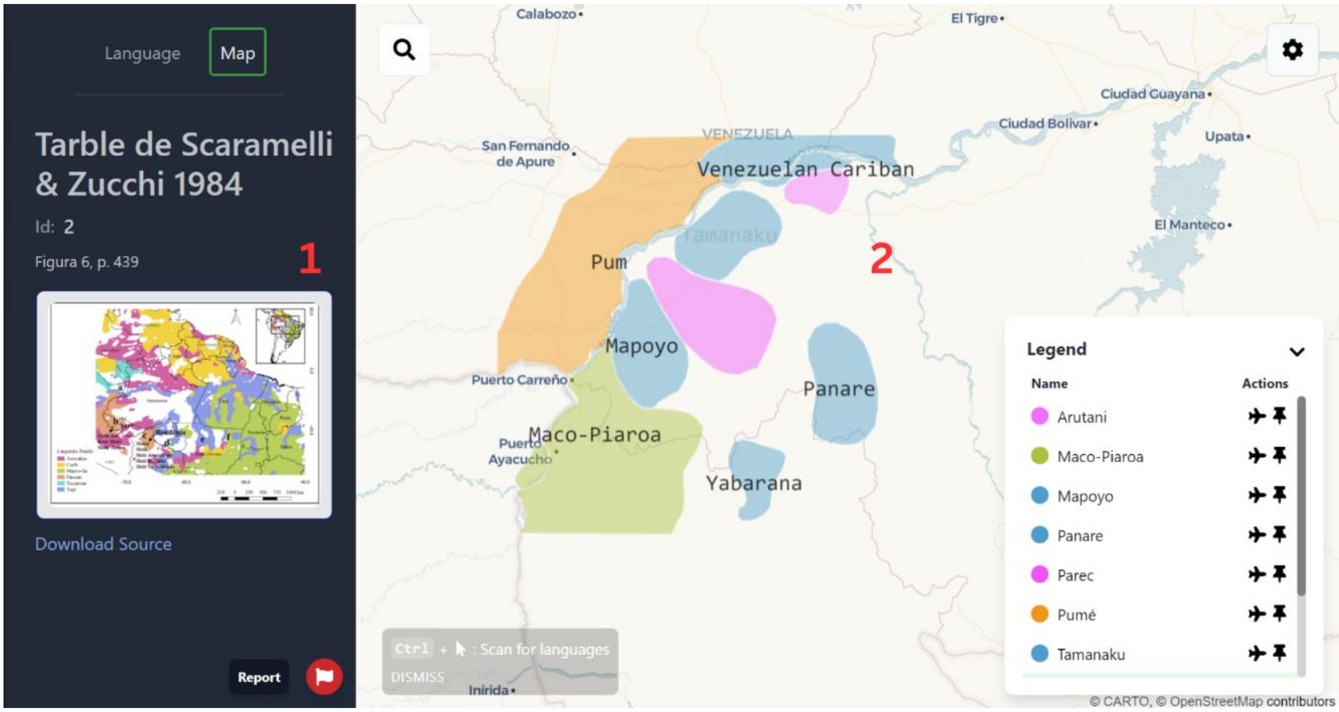


Figure 15 *- This Figure Presents the Final Glottography Prototype in the Map View, Highlighting New Features with Red Markers.*

Table 6 *- Feature Table for Figure 16 with Feature Titles and corresponding Description*

Number	Feature	Description
1	Source Image	A placeholder for the source image showing the origin of spatial information. It includes a download link which will become obsolete with the future initiation of a WFS (Web Feature Service).
2	Label with Collision Algorithm	Implements dynamic label adjustment with a collision detection algorithm to decide if a label should be displayed, varying with the map's zoom level.

Figure 16 illustrates the map view mode of the application, which remains largely unaltered from the language view. A notable feature is the embedding of a source map image directly in the map interface, providing users with the ability to ascertain the origin of spatial data. This is an essential aspect for validating data accuracy and integrity for scholarly use.

A second feature is the implementation of dynamic labeling. This feature improves languoid identification by adjusting label visibility based on the zoom level and available space, preventing overlap with neighboring labels and maintaining map legibility.

6.2 Task-Based Scenario Results

After presenting the prototype features, the focus shifts to the assessment results based on task-based scenarios. This analysis evaluates user task completion efficiency within the prototype, providing key usability insights. Importantly, this evaluation compares developer intentions (see previous chapter 5.1.) with user execution, indicating the prototype's effectiveness. Discrepancies between planned and actual usage inform enhancements for subsequent iterations, aiming for improved alignment and user experience optimization.

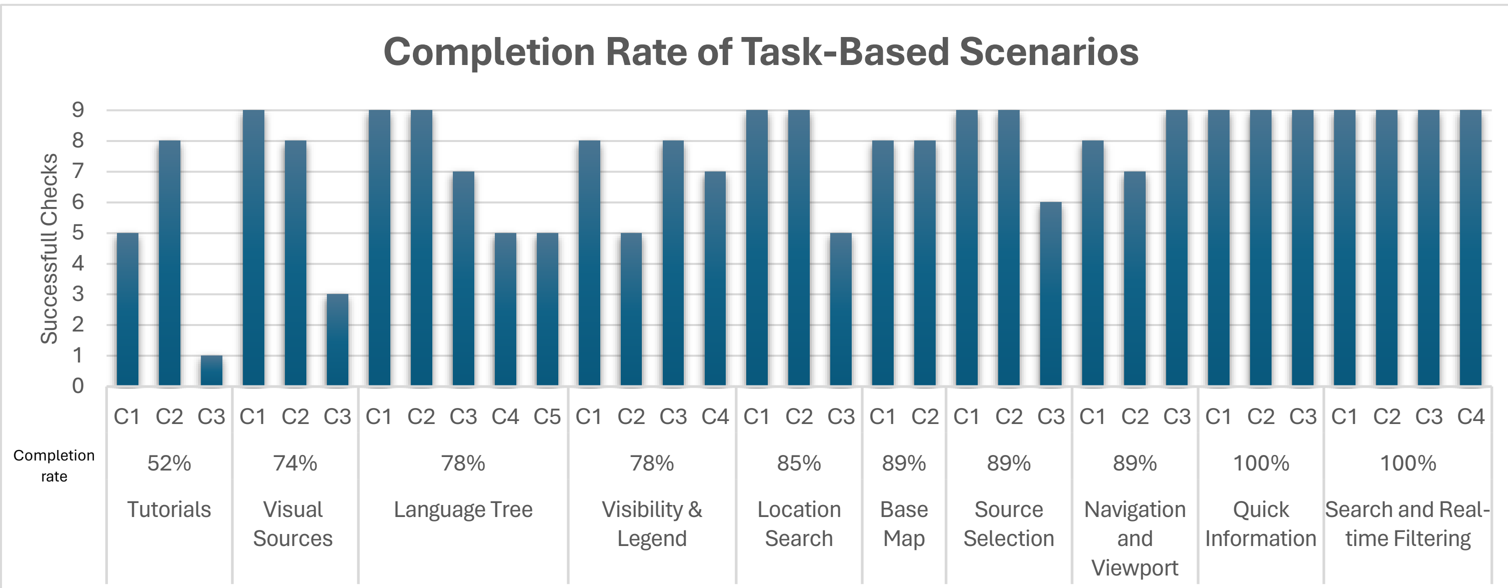


Figure 16 *- Evaluation of the Task-Based Scenarios with the Completion Rate of each Category and with the Individual Subtasks.*

The Figure 17 serves as an aggregated visual summary of the results from task-based scenarios, with each category on the x-axis corresponding to the bundled requirements category (see Subsection 5.3.1. & Appendix C). The columns within each category represent individual subtasks and indicate whether each subtask was successfully executed by the user. With nine participants in total, each subtask had the potential to be completed successfully nine times, equating to a 100% completion rate.

The categories are sequenced according to their aggregate completion rates, with "Tutorials" and "Visual Sources" on the left-hand side showing lower completion rates, whereas "Quick Information" and "Search and Real-time Filtering" on the right display full completion rates. This ordering visually emphasizes the variance in ease of use or understanding across different features.

The Task-Based Scenario Completion Graph suggests that every subtask was completed successfully by at least one participant, with overall task completion ranging from a minimum of 52% to a maximum of 100%. While this provides a broad measure of the prototype's performance across different tasks, detailed analysis of individual subtask results is necessary to discern the reasons behind the varying completion rates and to identify specific areas for improvement.

6.2.1 Sample Scenario Analysis 1

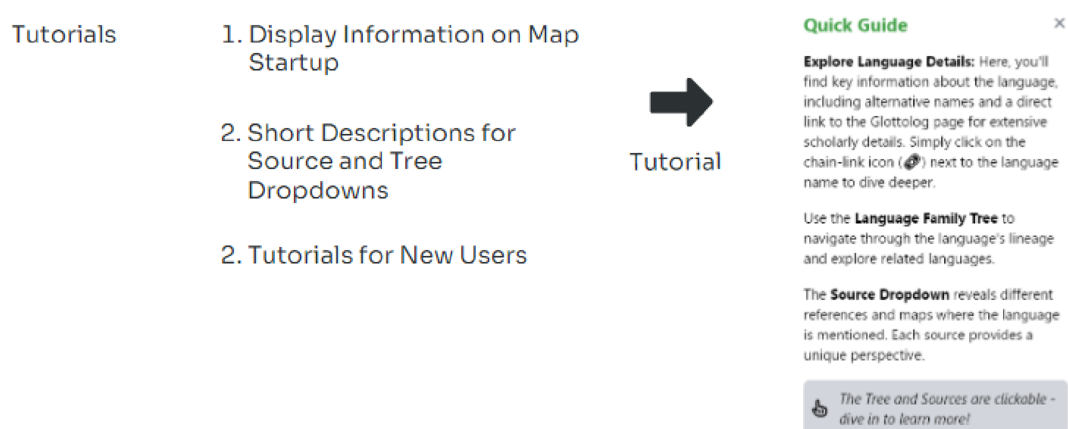
The "Tutorial" scenario's execution revealed that users were presented with a task deliberately left vague to mimic the varied approaches they might take when engaging with a new application. This approach also served as a test to see if users would naturally consult tutorial material or descriptions without explicit prompts. The task simply instructed users to find a language of interest, leaving them free to choose their method of discovery.

The subtasks focused on three behaviors: first, it noted whether users dived straight into utilizing the available tools to obtain language information or if they took preliminary steps to familiarize themselves with the interface. This involved scanning for interactive elements and reading the introductory content. Subsequently, once a language was selected, a dropdown guide appeared, providing an explanation of the information panel's components. Here, the assessment criteria included whether users took the initiative to read through the guide to gain a deeper insight into the interface's functionalities. The last observation was in regard to user interaction with the map interface after consulting the dropdown guide. It was found that almost all participants did not realize that the guide could be closed to unveil further language details. This also represents the low completion rate of the last subtask.

Tutorials and Descriptive Dropdowns

Provides initial guidance for new users through an informational panel and clear, informative dropdown guide to enhance starting experiences.

Task: "I'd like you to get acquainted with the map by finding information on a language you're interested in. How would you go about starting this task?"



Checklist:

- ✓ Participant browses initial map information and explains or demonstrates how they would initiate a search for specific language information.
- ✓ Participants reads map guide for orientation.
- ✓ Participant then proceeds to browse the language information component.

Figure 17 *- Task-Based Scenario Example. Tutorials and Descriptive Dropdown Category with short Description and the Checklist for the Assessment*

In an ideal interface design, user guidance would be minimal, with the architecture of the application itself intuitively leading users through tasks to successful completion. The low completion rate in the "Tutorial" scenario suggests a potential disconnect between the interface's design and user intuition, indicating areas that warrant attention and improvement in future design iterations. However, the ambiguity inherent in the task must be acknowledged. While a low completion rate can signal a design issue, it also doesn't capture the multifaceted nature of the task at hand. The deliberately vague nature of this task, as opposed to the more straightforward tasks in other scenarios, stems from the challenge of assessing whether users will seek out and read instructions without direct prompting. This type of behavior may not happen if the task's directive is overly descriptive.

While the completion rate is a crucial metric, it's equally important to highlight that a low completion rate can stem from multiple causes, yet generally, it suggests a misalignment between the developers' envisioned user workflow and actual user behavior. Such a discrepancy necessitates a reevaluation of the interface—particularly the tutorial setup—to better align with user needs.

In the instance at hand, the potential solutions seem straightforward yet impactful: resizing, repositioning, or more clearly indicating the close button could significantly improve success rates. These adjustments would likely enhance user interaction, bringing the developers' intentions and users' actions into closer harmony.

6.2.2 Sample Scenario Analysis 2

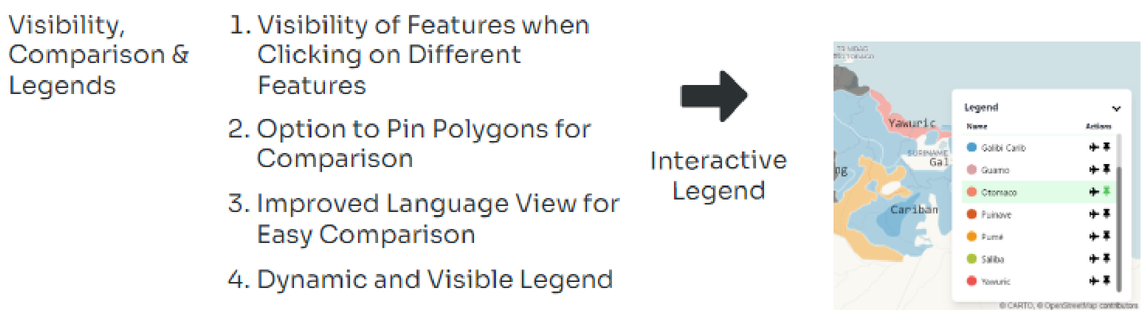
The "Visibility, Comparison, and Legends" scenario was crafted to assess users' proficiency with the map's comparative tools. Users were tasked to track and analyze linguistic data, a task that required them to intuitively navigate the interface's functionalities. The core of this evaluation lay in observing if participants could effectively use the pinning feature to compare language polygons and interpret the dynamic legend for spatial context. Unlike more directive tasks, this scenario aimed to uncover users' innate interaction strategies with the map's features.

Attention was given to whether users could identify and utilize the pinning function from both the map and the language information panel, and whether they understood the legend's role in correlating map elements with their data points. A critical part of the observation was to confirm that users recognized the highlighted languages as their active comparisons. The discovery that participants might miss the utility of certain tools, such as the pinning feature, suggests an opportunity to refine the interface to make these options more apparent and accessible.

Visibility, Comparison, and Legends

Focuses on maintaining clarity while interacting with multiple languages and improving comparative analysis tools.

Task: "You're analyzing linguistic data for a report and need to compare several languages. Determine how you might keep track of and compare your findings within the map."



Checklist:

- ✓ Participant pins language polygons to compare linguistic data through the legend.
- ✓ Participant pins language polygons to compare linguistic data through the language information panel.
- ✓ Participant reads and interprets the dynamic legend to understand what map features are displayed.
- ✓ Participant understands the highlighted languages represent pinned language polygons.

Figure 18 *- Task-Based Scenario Example. Visibility, Comparison, and Legends Category with short Description and the Checklist for the Assessment*

The insights gathered from this scenario point towards a need for a careful balance between a user-driven discovery process and the gentle nudges of interface guidance. The goal remains to enhance the interface to where its design naturally leads users through its features, allowing for a seamless comparative analysis experience. This scenario underscores the importance of designing for intuitive exploration, ensuring that even without explicit instructions, users can fully leverage the system's capabilities for their analytical tasks.

6.3 System Usability Test

The System Usability Test Results presented here reveal a spectrum of user experiences with the system in question. With an average SUS score of 75.21, the system is conventionally rated as 'good.' This score sits above the average threshold, suggesting that users generally find the system satisfactory in terms of usability.

The median score, which can often be a more robust measure as it is less affected by outliers, is slightly higher at 77.5, also described as 'good' and reinforcing the implication that the system has an above-average usability.

Notably, the range of scores indicates variability in user experience. The lowest SUS score recorded by a single participant is 65, which is still within the 'good' category, albeit it stands on the threshold and is considered 'below average' by conventional standards. This highlights that at least one user encountered notable usability issues.

Conversely, the highest score achieved is 92.5, which is categorized as 'best imaginable' or grade A. This score suggests that for some users, the system offers an exceptionally intuitive and user-friendly experience.

Overall, the SUS scores indicate that while most users seem to be satisfied with the system, there is room for improvement, especially considering the breadth between the lowest and highest scores. The goal for future iterations would be to address the issues causing lower scores while maintaining the elements that contribute to the high scores, aiming to elevate the overall user experience to consistently higher SUS ratings.

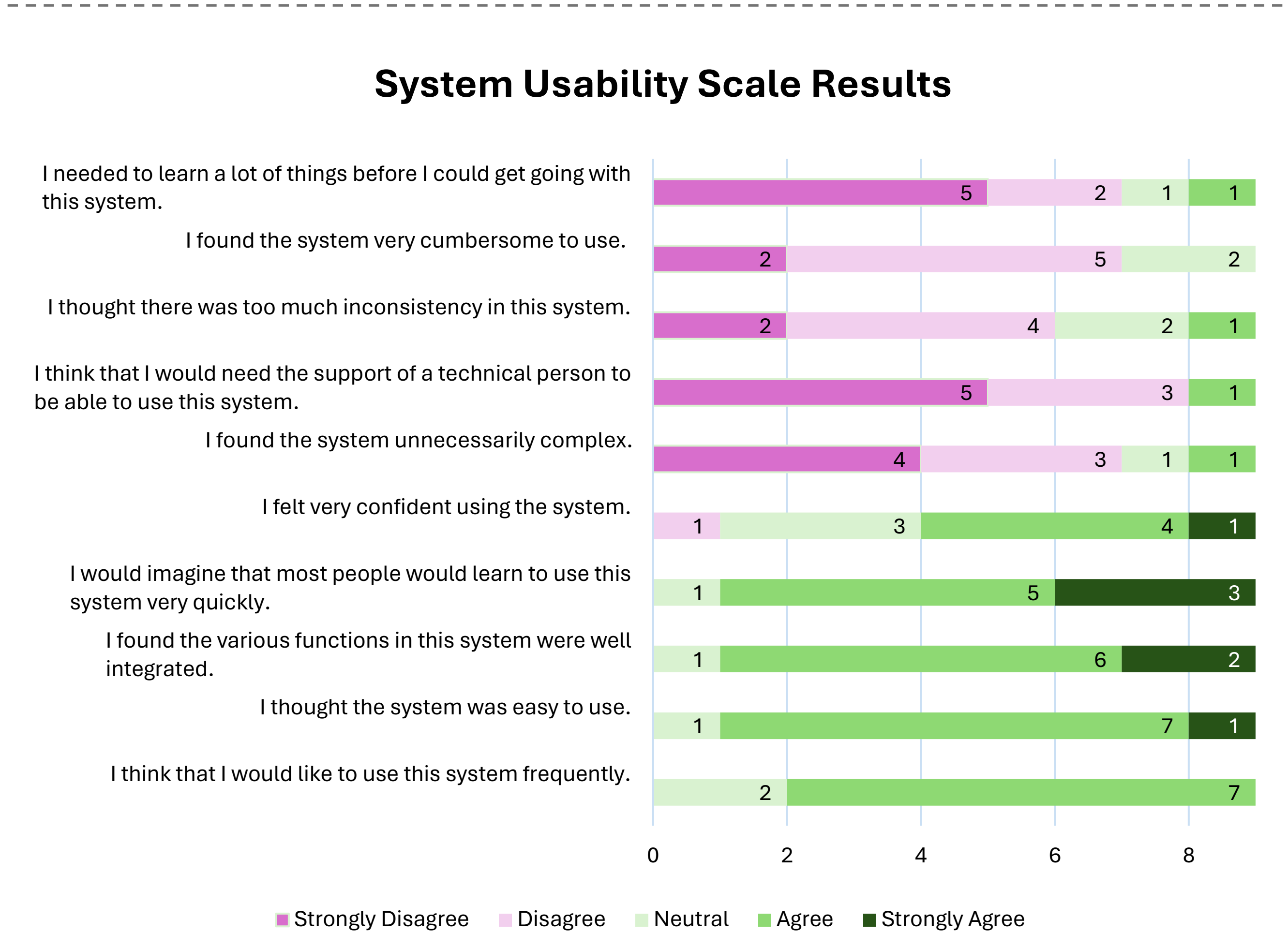


Figure 19 *- System Usability Scale Results with the 10 Questions. Note that the order of the questions is different here for visualization purposes and that in the original version there is questions go alternating from top and bottom,*

7. Discussion

This discussion centers on the prototype developed under the Glottography, aimed at addressing the need for a sophisticated tool capable of visualizing linguistic diversity globally, with a specific focus on South America. The prototype was crafted to fill the identified research gap: the absence of an accessible, user-centered web map service that accurately represents linguistic data through detailed polygonal mappings.

The analysis will evaluate whether UCD principles, along with direct user involvement and strategic design choices, have significantly improved the usability and satisfaction of the platform. Insights from this evaluation will guide potential improvements for future geovisualization platforms, ensuring they better meet the needs of the linguistic research community and advance the broader objectives of the Glottography Project.

7.1 Task-Based Scenarios Completion Analysis

The ‘Completion Rate of Task-Based Scenarios’ Graph presents a promising outcome, with high success rates in tasks such as 'Navigation and Viewport', 'Quick Information', and 'Search and Real-time Filtering'. These tasks, essential for the interactive exploration of spatial linguistic data, indicate that strategic design choices positively impacted the usability of the platform. The lower completion rates in 'Tutorials' and 'Visual Sources' suggest that initial user engagement with the platform may be encountering hurdles, possibly due to the learning curve or the presentation of instructional content.

In evaluating the 'Tutorials and Descriptive Dropdowns' scenario, it was observed that users frequently opted not to engage with the tutorial features. The users generally prefer to directly interact with the platform before seeking guidance. This behavior may not necessarily reflect an intuitive system design but rather a user tendency to explore independently until they encounter an obstacle. Acknowledging this user tendency, it may be beneficial for future design improvements to accommodate this explorative approach, perhaps by integrating tooltips or context-sensitive help that becomes available when the user's actions suggest they may need assistance.

The discovery that users turn to tutorials after facing difficulties rather than at the outset suggests two potential design considerations. First, it points to the importance of easily accessible, on-demand guidance that does not preemptively intrude on the user's exploration but is readily available when required. Second, it implies that while initial platform interactions may be intuitive, there could be complex functionalities that are not immediately obvious to users and could be made more accessible through navigational cues or assistance prompts.

The difficulty encountered by users in closing the tutorial guide to access more detailed language information underscores a need for optimization. Enhancing the visibility and accessibility of interface elements such as the close button is a critical design consideration, ensuring that users can seamlessly transition between tutorial guidance and independent exploration of the platform.

In assessing the 'Visibility, Comparison, and Legends' scenario, it became apparent that users encountered challenges with the platform's pinning feature used for comparing languages. The option to pin languages in two different areas, the information component and the legend, led to variability in how users utilized this function. Having more than one way to access pinning might offer flexibility but also seems to complicate the user experience, causing some to overlook this feature.

To improve consistency in user interaction, it may be necessary to streamline the pinning process. Simplification could be achieved by standardizing the visual language used to indicate the pinning option or by consolidating the pinning mechanisms into a single, intuitive control. This change aims to enhance the exploratory nature of the platform, making it easier for users to navigate and compare linguistic data effectively.

The dynamic legend, integral to comparing pinned languages, also warrants attention. Enhancing clarity in the legend when languages are selected for comparison could involve more explicit visual signals. Such improvements would support the platform’s primary function as a tool for exploration, ensuring that users can effortlessly track and contrast different linguistic data without the interface complexity that may differ from the user experience.

The insights from both scenarios contribute to the iterative design process. Future design iterations could focus on striking a balance between facilitating user discovery and providing clear pathways to the platform's full functionality. The goal is to create an environment where users can both feel confident exploring on their own and know that help is readily available and easy to utilize when required.

7.2 SUS Usability Test Results Analysis

The SUS Usability Test Results for the platform reveal a multifaceted user reception. Positive feedback on user confidence and the system's perceived ease of use contrasts with concerns over its complexity and the initial learning curve required. The mixed feedback on the system’s complexity and inconsistencies suggests that while UCD principles have guided the platform's development, the implementation of these principles may not fully meet user expectations in certain areas.

Given the context provided by research on the SUS, which indicates that scores can be effectively contextualized with adjective ratings to clarify the usability level, this analysis can be deepened (Bangor et al., 2009). For instance, if the platform's SUS score places it in a range that, according to the extended SUS research, aligns with adjectives such as "OK" or "acceptable," it prompts a reevaluation of what these ratings truly signify for the user experience. The designation of "OK" may superficially suggest adequacy, yet the detailed user feedback points towards specific usability challenges that differ from an optimal experience.

The division in user opinions, especially regarding the platform's need for pre-use learning and its perceived cumbersomeness, highlights critical areas for improvement. This disparity in user experiences underscores the necessity of refining the platform’s design to make it more intuitive and to reduce the reliance on external guidance or support. This could involve simplifying complex functionalities or enhancing the onboarding process to better accommodate users' initial interactions with the system.

The identification of these specific user experience challenges through the SUS results, interpreted within the broader framework of SUS research, emphasizes the importance of moving beyond average scores to address underlying issues. Recognizing that a mid-range SUS score may indicate a platform that is merely "OK" underlines the need for improvements to increase the user experience. This involves not just addressing the identified usability concerns but also rethinking the design approach to more closely align with user expectations and behaviors.

7.3 Synthesis of the Scenario and Usability Testing

While the overarching hypothesis that UCD enhances usability and user satisfaction holds, the feedback from users reveals a gap between theoretical UCD principles and their execution in practice. This discrepancy is particularly evident in two main areas: the standardization of tasks and the initial user-system interaction.

Standardization of Tasks: Users' experiences with the platform suggest that a more standardized approach to task design could improve usability. For instance, if users encounter two different mechanisms for a single function like pinning or searching, this can lead to confusion and a fragmented user experience. An example from the platform involves the pinning feature being accessible from multiple locations (information component and legend), which, while intended to offer flexibility, inadvertently complicates the user journey. Standardizing such tasks, by offering a single, intuitive method for execution, could streamline the user experience, making the system easier to navigate and reducing cognitive load.

Initial User-System Interaction: The initial interaction between the user and the system is crucial in setting the tone for the user experience. Feedback highlighted challenges in this area, with users often bypassing tutorials in favor of immediate interaction with the system. This behavior suggests a disconnect between the design of onboarding processes and user preferences for learning and exploration. For example, users expressed a desire to dive into the system and only seek help when encountering difficulties, indicating that a more dynamic, context-sensitive approach to tutorials and help features could be more effective. Instead of static, one-time tutorials, embedding contextual hints or pop-up guides that activate based on user actions could provide support exactly when needed, aligning with user tendencies to learn by doing.

7.4 Reflection on User-Centered Design Steps

7.4.1 Work Domain Analysis

The 'quick and dirty' approach adopted in the initial phase of the Glottography was particularly successful because it allowed for a responsive and user-focused development process (see Section 5.2.). By quickly implementing a basic interactive map, it was possible to immediately begin collecting valuable user feedback. This approach was highly efficient as it avoided the lengthy and often resource-intensive process of building a fully-featured prototype before understanding the users' needs. The rapid deployment of this first prototype meant that user feedback could inform the development from an early stage, ensuring that the most critical features were identified and prioritized (Maguire and Bevan, 2002).

The initial prototype facilitated an early test of the core functionalities that would be essential to the platform. For example, even with its simplicity, users could evaluate the initial prototype's fundamental interaction capabilities, such as updating languoids in the information panel which synchronizes map features. This immediate user interaction with the platform's basic functions provided quick, actionable insights that could be integrated into the development cycle much faster than in a traditional development process.

The structured interviews complemented the work domain analysis by providing a systematic method of gathering user feedback. The decision to organize discussions into three main categories ensured that the feedback was both focused and wide-ranging, covering necessary aspects of the platform's use. By approaching the interviews in this organized manner, specific user requirements could be drawn out, which informed subsequent design choices and helped prioritize development tasks.

Reflecting on the use of this initial prototype and the structured interviews, it’s clear that they provided an effective balance between rapid development and thorough user feedback collection. The prototype's simplicity ensured that user feedback was focused on the platform's essential aspects, while the interviews offered a depth of understanding that informed development decisions.

The limitations of this approach also became evident. The rapid development of the initial prototype risked missing more complex design issues that could arise later in the development process. It highlighted the need for a balance between speed and thoroughness. Future projects might benefit from incorporating iterative 'quick and dirty' prototyping phases with strategically placed in-depth analysis of certain aspects such as the phylogenetic tree, to capture both the high-level user requirements and more subtle design issues.

7.4.2 Conceptual Development & Prototype

The process of requirement analysis in the conceptual development and prototyping phase of the Glottography was integral to transforming user feedback into a coherent set of features for the platform. This systematic examination of user needs involved grouping them by feasibility, complexity, and necessity, which facilitated a more organized and efficient development process.

Requirement Grouping and Prioritization: Each user requirement was evaluated based on its technical feasibility, the complexity it would introduce to the system, and its necessity to the end-users. By categorizing these requirements, the team could create a prioritized roadmap for development. This structured approach ensured that resources were allocated effectively, focusing first on implementing essential features that offered the highest value to users. For example, the decision to deprioritize the tree view feature, which users found redundant, allowed for redirection of efforts towards more critical functionalities.

Development Guided by Grouped Requirements: The grouped requirements served as a blueprint for the subsequent design and development stages (also see Appendix C). By having a clear understanding of which features were most needed and feasible, the team could sequence their work to build out the platform incrementally, ensuring that each addition was aligned with user expectations and technical possibilities. This method of organizing requirements also highlighted the complexities associated with developing temporal analysis tools. Although historical linguists expressed a high need for such features, the limited availability of consistent historical data made a temporal feature implementation unfeasible in this iteration of the platform. This realization prompted a reassessment of priorities, where the team acknowledged the challenge and deferred the development of temporal tools until more comprehensive data could be secured or alternative solutions could be explored in future updates.

User Stories as a Development Compass: A crucial component of this phase was the creation of user stories (see Appendix B). These narratives played a vital role in consolidating the understanding of user requirements. They provided a mental map for the developer, enhancing the clarity of how the final prototype would be used. User stories acted as touchstones for functionality and design, ensuring that each feature was developed with a clear purpose and end-user scenario in mind. For example, a user story about a comparative linguist seeking to analyze language overlaps in multilingual regions helped prioritize the development of an intuitive pinning feature on the map interface.

During the development of this prototype, the engagement of a diverse range of stakeholders was crucial. By consulting with experts in fields such as historical linguistics, dialectology, sociolinguistics, and geospatial interdisciplinary research, the platform began to take shape in a manner that accommodated a broad spectrum of linguistic research needs. During the work domain analysis, the initial prototype's tree view feature, designed to depict language family relationships, was met with feedback indicating that it was not as valuable as anticipated. Users noted that such hierarchical information was readily available from other sources or did not align with their primary interests in the platform, emphasizing the importance of focusing on unique features that would offer new insights or functionality not found elsewhere. The input from historical linguists highlighted a desire for temporal analysis tools that would allow tracking language change over time. The feasibility of this was limited by the availability of historical linguistic data, which is often sparse or inconsistently recorded, presenting a significant challenge in the development of such tools within the platform.

7.4.3 Interaction & Usability Studies

The interaction and usability studies chapter of the Glottography is important in assessing the effectiveness of the implemented features and understanding user satisfaction. This subsection outlines the significance of the task-based scenarios and the usability test in gauging the platform's performance from a user-centered perspective.

The studies conducted as part of the thesis were pivotal in assessing the graphical user interface's effectiveness and overall user satisfaction. The engagement of nine diverse participants, encompassing a range of backgrounds, ages, and domain expertise, was strategic. Research suggests that conducting usability tests with only five users can reveal about 80% of interface problems 70% of the time (Faulkner, 2003). Expanding the test group to ten participants increases the minimum problem discovery rate to 80%, confirming the effectiveness of this sample size in gathering significant usability insights. Additionally, a diverse participant group further enables an assessment of the graphical user interface's (GUI) usability across various user demographics.

Task-based scenarios in the Glottography were instrumental because they are work-oriented design objects. They allow for directly validating the translation of user requirements into practical application features, since they emulate the kind of work users want to do on the system (Carrol, 2003). Each scenario was crafted to mimic real-world tasks that users might perform on the platform, thus providing an authentic context for testing. The detailed breakdown of each task into subtasks allowed for pinpointing specific areas of the platform that worked well or needed refinement.

For example, one scenario involved users finding and utilizing visual sources related to linguistic data. The subtasks included locating the visual source tool, activating the tool to view a source, and then closing the tool to return to the main interface. The absence of a clear function to close the enlarged source image highlighted a specific usability gap. By isolating this issue to a particular subtask, the focus could be set on refining this interaction without the need to reconsider the entire visual source functionality. This granularity not only made problem-solving more efficient but also helped maintain the integrity and value of the overall feature set. This methodical breakdown of tasks into actionable subtasks is particularly valuable in iterative testing and development, as it allows developers to make precise adjustments based on well-defined user interactions and feedback.

Task completion times were not individually measured during these studies. However, there was a broader time constraint applied, with each interview session capped at a maximum of one hour. This approach balanced the need for comprehensive qualitative feedback with the practical limitations of participant engagement time. Opting not to measure individual task completion times allowed participants to focus on providing in-depth feedback without the pressure of a ticking clock, fostering a more thorough exploration of the interface's usability and potential areas for improvement. In future iterations, timing specific tasks may be incorporated to better assess efficiency and allow for comparisons of similar projects.

The usability test provided a complementary broad assessment of the platform's usability and user satisfaction. As a standardized tool, the SUS gives an overall usability score that can be compared against other software products, offering an external benchmark for evaluating the platform's performance. The general nature of the SUS feedback is its strength, as it captures a holistic view of the user experience. For the Glottography, applying the SUS test helped to contextualize the user satisfaction level in relation to other applications. It provided a snapshot of how intuitive and satisfactory the platform was perceived by its users, which is crucial for identifying whether the platform meets the basic usability expectations.

The SUS results can be helpful in guiding broader strategic improvements. For instance, if the SUS indicates lower scores in areas related to the ease of learning how to use the platform, this could suggest a need for improved tutorials or a more intuitive interface design. Conversely, high SUS scores can validate the effectiveness of the current design and functionality, endorsing the current developmental path and encouraging further enhancement in similar directions.

Integrating the detailed insights from task-based scenarios with the broad evaluative perspective provided by the SUS offers a robust framework for ongoing development. This approach ensures that while specific usability issues are systematically addressed, the overall user experience is progressively refined based on empirical feedback and performance evaluations. The concurrent use of task-based scenarios and the usability test supports a balanced development strategy that emphasizes both targeted feature improvements and broad enhancements to user interaction. This approach guarantees that the platform functions effectively while also adapting to user needs and expectations, ensuring its robustness and user-friendliness within its specific application context. Although usability testing is a valuable tool, it does not ensure that all potential usability issues will be identified (Lewis, 2012). Its effectiveness, demonstrated by the positive results in this thesis, does not negate the necessity for a cautious interpretation of these outcomes. Usability testing is inherently imperfect, and it is essential to critically assess its results. A thorough understanding of its capabilities, limitations, and best practices is essential to apply it effectively and to make the most of its benefits while acknowledging its constraints.

Regarding the comparison of GUI usability across different systems, it is recognized that effectiveness and efficiency measures may vary significantly, complicating direct comparisons. While it is challenging to definitively state that one system is more usable than another due to these variable metrics, high-level subjective assessments of usability can still enable meaningful comparisons (Bangor et al., 2008). This approach enables the evaluation of how the GUI of the Glottography project compares to user expectations and established usability standards, even without direct comparisons between systems.

7.5 Discussion for Linguistic Database GUI

The development of the GUI for the linguistic database has laid a robust foundation for a simplified data presentation, enhancing user navigation, and overall ease of use. The current implementation, particularly the method of displaying language polygons one at a time coupled with the ability to pin and compare, has been well-received by users. This reception provides a strong basis for considering further enhancements and exploring additional aspects of the cartographic visualization.

A key aspect of this research is the potential transferability of its outcomes. The design strategies and iterative development processes employed, including a complete cycle demonstrated in this thesis, suggest significant potential for future projects. Even though this project encompassed just one full iterative cycle, it showcased the substantial benefits of this approach, which can be applied to other geospatial visualization projects beyond linguistics. For instance, these methodologies could be adapted for environmental studies, urban planning, or any field that requires the integration of complex spatial data into user-friendly interfaces. This adaptability highlights the broader applicability of the research findings, making them valuable across various domains that deal with geographic information.

The project's geographic nature is evident in its focus on spatial representations of linguistic data. By mapping languages and allowing for interactive engagements with these maps, the GUI serves not only as a tool for linguistic research but also as a critical resource for exploring geographical layers of linguistic distribution. This geographical component is essential, as it provides unique insights into how languages are distributed across physical spaces, potentially reflecting historical migrations, cultural boundaries, and socio-political divisions.

While the platform currently supports basic functionalities for viewing and comparing linguistic data, there is room for expansion. It is crucial to maintain the platform's primary role as a tool for exploration, ensuring that it remains accessible and straightforward for users primarily interested in navigating and visualizing linguistic data. Although the GUI is designed for intuitive use, it functions as an entry point for users to interact with linguistic diversity through customizable and interactive visual representations.

Additionally, expanding the GUI to include more customizable visualization options might cater to a wider range of user preferences and research needs. Enabling users to overlay additional datasets or create their own visual representations could make the platform more versatile and useful.

8. Conclusion and Future Work

8.1 Contributions

Assessing the Glottography reveals a detailed view of the UCD approach's effectiveness, illustrating both its successes and limitations. Initially, the project hypothesized that direct user involvement and strategic design decisions would significantly enhance usability and the effectiveness of spatial linguistic data presentation. While findings from usability tests, task-based scenarios, and comprehensive analyses of the UCD process largely support this hypothesis, they also highlight several critical areas for improvement.

The usability tests and task-based scenarios demonstrated that certain design choices, particularly those aimed at simplifying navigation and clarifying data presentation, were effective and resonated well with users. These elements undoubtedly facilitated deeper user engagement and improved understanding of linguistic data, with many users navigating the platform intuitively, which suggests a successful integration of key UCD principles. However, these positive outcomes should not obscure the fact that the implementation of UCD did not fully meet expectations across all dimensions of the project.

The iterative design process, a core component of UCD, unveiled significant ongoing needs for refinement. Analyzing each phase of the UCD process uncovered various deficiencies, revealing that the project sometimes struggled to anticipate and meet complex user needs. This process highlighted that while responsiveness to user feedback is essential, it also led to numerous challenges. Essential features were occasionally postponed or scaled back due to practical constraints such as feasibility issues or resource limitations, which disrupted the planned development trajectory.

In conclusion, although not every anticipated feature was implemented, such as temporal analysis tools constrained by data availability, the primary objectives to enhance usability and user satisfaction were broadly met. Nevertheless, the project's experience underscores that UCD, while beneficial, requires careful and adaptive application. It's a dynamic approach that must flexibly respond not only to user feedback but also to ongoing project constraints and evolving requirements. This critical reflection on the Glottography provides valuable lessons on the application of UCD in complex systems, offering insights that could refine future projects aiming to meld advanced technical solutions with user-centric design principles. This balanced perspective ensures a more realistic appraisal of UCD's impact and its potential in future technological and research endeavors.

8.2 Limitations

A key limitation of the current platform is the limited availability and completeness of data, particularly in terms of speaker counts and temporal dynamics, which are crucial for providing users with comprehensive linguistic context. Enhancing the scope and quality of these datasets would significantly benefit user engagement and analysis capabilities. Additionally, the technical execution of the platform is limited by the author's current level of expertise. The implementation of more complex functionalities, as envisioned, was restricted, pointing to the need for more skilled developers or a larger team. Expanding the development team would not only improve the platform's current features but also enable the integration of advanced tools and analytical capabilities, thus elevating the overall functionality and user experience.

Moreover, the elicitation of user requirements and their subsequent translation into platform features are subject to developer bias. This bias arises from the developer's interpretation of feedback and decisions about feature prioritization and implementation, which may not accurately reflect the users' needs or preferences. Such biases could affect the reproducibility of the research and the generalizability of the platform across different user groups and use cases.

Testing variability also presents a significant limitation. The platform's evaluation through task-based scenarios inherently leads to varied outcomes depending on what specific functionalities are tested and how they are weighted by different researchers, based on their interests and perspectives. This subjective approach to testing can result in inconsistent findings and may not comprehensively address all relevant aspects of usability and functionality.

These limitations underscore the need for a methodologically robust approach in future developments. Addressing these challenges requires systematic enhancements in data collection, a broadening of developer expertise, and the adoption of more standardized testing protocols to ensure that subsequent iterations of the platform offer improvements that are both meaningful and user-focused.

8.3 Future Research Recommendations

Enhancing the GUI to support the overlay of additional datasets, execute simple transformations, and customize visual representations could significantly boost the exploratory capabilities of the platform. This functionality would enable users to adapt their interactions based on specific research interests or curiosities, enriching their engagement without complicating the core user experience. At present, the platform facilitates basic data manipulations and visual customizations while deliberately omitting complex data management or detailed statistical analysis tools. This strategic decision maintains a streamlined interface that prioritizes exploration, while guiding users who require in-depth analyses to download the data for use with more specialized GIS software or other analysis tools.

Future research should also consider evaluating and refining the visual styles used for language mapping within the platform. There remains a noticeable gap in research specifically addressing the effectiveness of various mapping styles for language data. Investigating how to maintain clarity and focus when users switch between viewing individual languages, related language groups, or specific source maps could lead to actionable recommendations or even the implementation of new WebGIS styles tailored for language mapping. This focus on optimizing language map styles could greatly improve how linguistic ranges are visualized, potentially incorporating expert knowledge to offer more detailed and nuanced representations. These advancements would accommodate a broader spectrum of scholarly needs and potentially foster deeper insights into linguistic geographies, thereby extending the platform's utility and impact in linguistic research.

References

Bangor, A., Kortum, P., Miller, J., 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale 4.

Bangor, A., Kortum, P.T., Miller, J.T., 2008. An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction 24, 574–594. https://doi.org/10.1080/10447310802205776

Bano, M., Zowghi, D., Ferrari, A., Spoletini, P., Donati, B., 2019. Teaching requirements elicitation interviews: an empirical study of learning from mistakes. Requirements Eng 24, 259–289. https://doi.org/10.1007/s00766-019-00313-0

Berander, P., Andrews, A., 2005. Requirements Prioritization, in: Aurum, A., Wohlin, C. (Eds.), Engineering and Managing Software Requirements. Springer, Berlin, Heidelberg, pp. 69–94. https://doi.org/10.1007/3-540-28244-0_4

Brooke, john, 1996. SUS: A "Quick and Dirty" Usability Scale, in: Usability Evaluation In Industry. CRC Press.

Campbell, L., 2007. Glossary of historical linguistics. Edinburgh University Press.

Campbell, Lyle, and Verónica Grondona. *The Indigenous Languages of South America: A Comprehensive Guide*. 2. Walter de Gruyter, 2012. https://doi.org/10.1515/9783110258035.

Carrol, J.M., 2003. Scenario-Based Desgin, in: Making Use: Scenario-Based Design of Human-Computer Interactions. MIT Press, pp. 45–70.

Cartwright, W., Peterson, M.P., 2007. Multimedia Cartography, in: Cartwright, W., Peterson, M.P., Gartner, G. (Eds.), Multimedia Cartography. Springer, Berlin, Heidelberg, pp. 1–10. https://doi.org/10.1007/978-3-540-36651-5_1

Çöltekin, A., Bleisch, S., Andrienko, G., Dykes, J., 2017. Persistent challenges in geovisualization – a community perspective. International Journal of Cartography 3, 115–139. https://doi.org/10.1080/23729333.2017.1302910

Çöltekin, A., Janetzko, H., Fabrikant, S.I., 2018. Geovisualization. Geographic Information Science 2018, online.

Daurio, M., Turin, M., 2019. "Langscapes" and language borders: Linguistic boundary-making in northern South Asia 10, 21–42.

Dryer, M.S., Haspelmath, M. (Eds.), 2013. WALS online (v2020.3). https://doi.org/10.5281/zenodo.7385533

Dykes, J., Lloyd, D., Radburn, R., 2007. Understanding geovisualization users and their requirements: a user-centred approach, in: GIS Research UK 15th Annual Conference (GISRUK 2007). Maynooth, Ireland.

Dykes, J., MacEachren, A.M., Kraak, M.-J., 2005. Introduction Exploring Geovisualization, in: Dykes, J., MacEachren, A.M., Kraak, M.-J. (Eds.), Exploring Geovisualization, International Cartographic Association. Elsevier, Oxford, pp. 1–19. https://doi.org/10.1016/B978-008044531-1/50419-X

Eberhard, D.M., Simons, G.F., Fennig, C.D., 2023. Ethnologue: Languages of the world, 26th ed. SIL International, Dallas.

Edsall, R., Andrienko, G., Andrienko, N., Buttenfield, B., 2008. Interactive maps for exploring spatial data. Manual of geographic information systems 837–858.

Faulkner, L., 2003. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers 35, 379–383. https://doi.org/10.3758/BF03195514

Forkel, R., Hammarström, H., 2022. Glottocodes: Identifiers linking families, languages and dialects to comprehensive reference information. Semantic Web 13, 917–924. https://doi.org/10.3233/SW-212843

Forkel, R., List, J.-M., Greenhill, S.J., Rzymski, C., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G.A., Gray, R.D., 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. Sci Data 5, 180205. https://doi.org/10.1038/sdata.2018.205

Fox, J.E., 2015. The science of usability testing, in: Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC, USA. pp. 1–3.

Goguen, J.A., Linde, C., 1993. Techniques for requirements elicitation, in: [1993] Proceedings of the IEEE International Symposium on Requirements Engineering. Presented at the [1993] Proceedings of the IEEE International Symposium on Requirements Engineering, pp. 152–164. https://doi.org/10.1109/ISRE.1993.324822

Good, J., Hendryx-Parker, C., 2006. Modeling contested categorization in linguistic databases, in: Proceedings of the EMELD 2006 Workshop on Digital Language Documentation: Tools and Standards: The State of the Art. Lansing, Michigan.

Granlund, Å., Lafrenière, D., Carr, D.A., 2001. A Pattern-Supported Approach to the User Interface Design Process.

Griffin, A.L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C.R., Bravo, J.V.M., Fabrikant, S.I., Bleisch, S., Yamada, M., Picanço, P., 2017. Designing across map use contexts: a research agenda. International Journal of Cartography 3, 90–114. https://doi.org/10.1080/23729333.2017.1315988

Hammarström, H., Forkel, R., Haspelmath, M., Bank, S., 2022. Glottolog 4.7. Leipzig. https://doi.org/10.5281/zenodo.7398962

Hasani, L.M., Sensuse, D.I., Suryono, R.R., 2020. User-Centered Design of e-Learning User Interfaces: A Survey of the Practices.

Haynie, H., 2014. Geography and Spatial Analysis in Historical Linguistics. Language and Linguistics Compass 8. https://doi.org/10.1111/lnc3.12087

Haynie, H.J., Gavin, M.C., 2019. Modern Language Range Mapping for the Study of Language Diversity. https://doi.org/10.31235/osf.io/9fu7g

Hoch, S., Hayes, J., 2010. Geolinguistics: The Incorporation of Geographic Information Systems and Science.

Holtzblatt, K., Wendell, J.B., Wood, S., 2004. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design. Elsevier.

International Organization for Standardization, 2019. Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems (ISO standard No. ISO 9241-210:2019). International Organization for Standardization, Geneva, Switzerland.

International Organization for Standardization, 1999. Human-centred design processes for interactive systems (ISO standard No. ISO 13407:1999). International Organization for Standardization, Geneva, Switzerland.

Kramers, E.R., 2008. Interaction with Maps on the Internet – A User Centred Design Approach for The Atlas of Canada. The Cartographic Journal 45, 98–107. https://doi.org/DOI: 10.1179/174327708X305094

Lewis, J.R., 2012. Usability Testing, in: Handbook of Human Factors and Ergonomics. John Wiley & Sons, Ltd, pp. 1267–1312. https://doi.org/10.1002/9781118131350.ch46

Lucassen, G., Dalpiaz, F., Werf, J.M.E.M. van der, Brinkkemper, S., 2016. The Use and Effectiveness of User Stories in Practice, in: Daneva, M., Pastor, O. (Eds.), Requirements Engineering: Foundation for Software Quality, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 205–222. https://doi.org/10.1007/978-3-319-30282-9_14

Luebbering, C.R., 2011. The Cartographic Representation of Language: Understanding language map construction and visualizing language diversity.

Luebbering, C.R., Kolivras, K.N., Prisley, S.P., 2013. Visualizing Linguistic Diversity Through Cartography and GIS. The Professional Geographer 65, 580–593. https://doi.org/10.1080/00330124.2013.825517

Maguire, M., Bevan, N., 2002. User Requirements Analysis, in: Hammond, J., Gross, T., Wesson, J. (Eds.), Usability: Gaining a Competitive Edge. Springer US, Boston, MA, pp. 133–148. https://doi.org/10.1007/978-0-387-35610-5_9

Mao, J.-Y., Vredenburg, K., Smith, P.W., Carey, T., 2005. The state of user-centered design practice. Commun. ACM 48, 105–109. https://doi.org/10.1145/1047671.1047677

Mao, J.-Y., Vredenburg, K., Smith, P.W., Carey, T., 2001. User-centered design methods in practice: a survey of the state of the art, in: Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research. p. 12.

Marsh, S.L., Haklay, M., 2010. Evaluation and deployment, in: Haklay, M. (Ed.), Interacting with Geospatial Technologies. Wiley-Blackwell, West Sussex, UK, pp. 199–221.

McNew, G., Derungs, C., Moran, S., 2018. Towards faithfully visualizing global linguistic diversity, in: Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., Tokunaga, T. (Eds.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Presented at the LREC 2018, European Language Resources Association (ELRA), Miyazaki, Japan.

Nakić, J., Kosović, I.N., Franić, A., 2022. User-centered design as a method for engaging users in the development of geovisualization: A use case of temperature visualization. Applied Sciences 12. https://doi.org/10.3390/app12178754

Nelson, J.K., MacEachren, A.M., 2020. User-centered Design and Evaluation of a Geovisualization Application Leveraging Aggregated Quantified-Self Data. Cartographic Perspectives 7–31. https://doi.org/10.14714/CP96.1631

Nivala, A.-M., Brewster, S., Sarjakoski, L., 2008. Usability Evaluation of Web Mapping Sites. The Cartographic Journal 45. https://doi.org/10.1179/174327708X305120

Norman, D.A., Draper, S.W., 1986. User Centered System Design; New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., USA.

Rabanus, S., 2019. Language Mapping Worldwide: Methods and Traditions, in: Brunn, S.D., Kehrein, R. (Eds.), Handbook of the Changing World Language Map. Springer International Publishing, Cham, pp. 1–27. https://doi.org/10.1007/978-3-319-73400-2_151-1

Rantanen, T., Tolvanen, H., Roose, M., Ylikoski, J., Vesakoski, O., 2022. Best practices for spatial language data harmonization, sharing and map creation—A case study of Uralic. PLOS ONE 17, e0269648. https://doi.org/10.1371/journal.pone.0269648

Robinson, A.C., Chen, J., Lengerich, E.J., Meyer, H.G., MacEachren, A.M., 2005. Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. Cartogr Geogr Inf Sci 32, 243–255. https://doi.org/10.1559/152304005775194700

Robinson, A.C., Demšar, U., Moore, A.B., Buckley, A., Jiang, B., Field, K., Kraak, M.-J., Camboim, S.P., Sluter, C.R., 2017. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. International Journal of Cartography 3, 32–60. https://doi.org/10.1080/23729333.2016.1278151

Roose, M., Nylén, T., Tolvanen, H., Vesakoski, O., 2021. User-Centred Design of Multidisciplinary Spatial Data Platforms for Human-History Research. ISPRS International Journal of Geo-Information 10, 467. https://doi.org/10.3390/ijgi10070467

Roth, R., 2019. How do user-centered design studies contribute to cartography? Geografie 124, 133–161. https://doi.org/10.37040/geografie2019124020133

Roth, R.E., Çöltekin, A., Delazari, L., Filho, H.F., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., Ooms, K., van Elzakker, C.P.J.M., 2017. User studies in cartography: opportunities for empirical research on interactive maps and visualizations. International Journal of Cartography 3, 61–89. https://doi.org/10.1080/23729333.2017.1288534

Sauro, J., 2011. SUStisfied? Little-Known System Usability Scale Facts - User Experience. User Experience Magazine 10.

Sharp, H., Rogers, Y., Preece, J., 2007. Design, prototyping and construction, in: Interaction Design. Beyond Human-Computer Interaction. Wiley, John and Sons Ltd.

Sutcliffe, A.G., Maiden, N.A.M., Minocha, S., Manuel, D., 1998. Supporting scenario-based requirements engineering. IEEE Transactions on Software Engineering 24, 1072–1088. https://doi.org/10.1109/32.738340

Tsou, M.-H., 2011. Revisiting web cartography in the united states: the rise of user-centered design. Cartography and Geographic Information Science 38, 250–257. https://doi.org/10.1559/15230406382250

Van Gijn, R., Hammarström, H., Van De Kerke, S., Krasnoukhova, O., Muysken, P., 2017. Linguistic Areas, Linguistic Convergence and River Systems in South America, in: Hickey, R. (Ed.), The Cambridge Handbook of Areal Linguistics. Cambridge University Press, pp. 964–996. https://doi.org/10.1017/9781107279872.034

Vredenburg, K., Mao, J.-Y., Smith, P.W., Carey, T., 2002. A survey of user-centered design practice, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '02. Association for Computing Machinery, New York, NY, USA, pp. 471–478. https://doi.org/10.1145/503376.503460

Wong, D.W.S., 2004. The Modifiable Areal Unit Problem (MAUP), in: Janelle, D.G., Warf, B., Hansen, K. (Eds.), WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the 100th Anniversary of the Association of American Geographers 1904–2004. Springer Netherlands, Dordrecht, pp. 571–575. https://doi.org/10.1007/978-1-4020-2352-1_93

Appendix A – User Elicitation Interview Guideline

General

1. Introduction and warm-up:
    - Can you tell me a bit about your background and your interest in languages and maps?
    - Have you used any language visualization tools or applications before? If so, could you describe your experience with them?
2. General understanding of language visualization:
    - How familiar are you with language visualization tools or applications?
    - Could you share any experiences you've had with existing language maps or platforms?
    - When you interact with maps, how do you typically find information about languages?

Application specific

3. Exploration of current prototype:
    - What are your initial impressions of the prototype I've shared with you? Any thoughts or feedback?
    - Did you find the map, pop-up selection, and sidebar intuitive to use?
    - How clear and understandable were the visual elements, as the language ranges and how they are depicted?
4. Desired language information:
    - When selecting a language, what specific details or attributes would you like to see?
    - Are there any particular data points that you think would be most relevant and useful to you?
    - Would you be interested in information about language speakers, dialects, or historical context?

    Possible follow-up questions:

    - Can you give me an example of a specific data point or attribute you would like to have access to?
    - How would you prioritize different pieces of language information if you had to choose?

5. Different views and perspectives:
    - Are there alternative ways you would prefer to visualize language ranges?
    - How do you feel about using color gradients, overlays, or other visual cues to represent language ranges?
    - Would you like to have interactive features, such as toggling between different views?

    Possible follow-up questions:
    - Can you describe a specific map style or theme that you find visually appealing or effective?
    - Are there any particular interactive features that you believe would enhance your experience?
6. Language family tree:
    - How important is it for you to see the hierarchical relationships between languages?
    - Would you like the language family tree to be displayed separately or integrated with the language ranges?

    Possible follow-up questions:
    - What level of detail would you expect to see in the language family tree? For example, would you like to see subfamilies or only the main language families?
    - How would you prefer the language family tree to be represented visually? Are there any existing examples you find appealing?
7. Spatial references and sources:
    - Would it be valuable to you to have references for languages, such as scientific literature / journals or atlases showcased in the map context?

    Possible follow-up questions:
    - Can you provide an example of how spatial references could enhance your understanding or use of the language visualization?
    - Would you like to see the sources or references displayed directly in the application or as additional information that can be accessed separately?
8. User interaction and customization:
    - Are there any specific user interaction options or customization features you would like to have?
    - How important is the ability to filter or search for specific languages or language families?
    - Are there any other interactive features you would find useful?

    Possible follow-up questions:

- Can you describe a scenario where filtering or searching for specific languages or language families would be beneficial to you?
- Are there any customization options or features from other applications that you find particularly useful or appealing?

Finalizing

9. Closing and feedback:
    - Is there any additional feedback, suggestions, or concerns you would like to share?
    - Thank you for your time and input. Is there anything else you would like to add before we conclude the interview?

Appendix B – User Profiles and User Stories

Participant 1

Goals and Objectives:

- To explore and study the distribution and migration of languages across different regions and time periods.
- To investigate the relationships between languages, focusing on their historical development and genealogy.
- To analyze language data in a visual and interactive manner to gain deeper insights into language evolution and temporal context.
- To use the application as a valuable resource for academic research, language documentation, and teaching materials.

Needs and Preferences:

- A user-friendly interface that allows easy exploration of language data on the map.
- The ability to filter and visualize language data for different epochs or time periods, enabling historical analysis.
- Clear and detailed language information presented in the sidebar, including language family trees and historical references.
- Customization options to adjust the map view and tailor the visualization to suit specific research interests.
- Access to reliable sources and references that support the linguistic data displayed on the map.

Challenges:

- Dealing with vast amounts of language data and ensuring the accuracy and completeness of historical information.
- Navigating complex language relationships and understanding the intricacies of language evolution.
- Balancing the need for in-depth historical context with a user-friendly and visually appealing interface.

Scenario:

The participant is currently engaged in a research project focused on reconstructing the migration patterns of a specific language family across various epochs. To investigate how this language family originated and evolved over time, the participant utilizes a language visualization application, setting filters to display data relevant to the chosen time periods. The application presents the linguistic ranges for each epoch, enabling the participant to observe the expansion, contraction, and interactions of the languages with neighboring groups. The feature that outlines the family tree of languages is particularly useful, allowing the participant to trace historical relationships and deepen their understanding of linguistic development. The temporal filters and historical context provided by the application are essential tools for the participant, aiding in the identification of significant patterns and yielding valuable insights into the language data.

Key Takeaways:

Participant 1, an experienced professor specializing in historical linguistics, primarily uses the application to examine language data through a temporal perspective to understand language evolution and migration patterns. Enabling the participant to filter and visualize language data across different epochs, coupled with access to explicit historical references, significantly augments their research capabilities and supports their research.

User stories:

Explore Language Data with Temporal Context

User Story: As a linguistic professor, I want to filter the language data on the map for different epochs so that I can analyze the temporal context of language distribution and migration.

Analyze Language Relationships and Family Trees

User Story: As a linguistic researcher, I want access to language family trees displayed in the sidebar so that I can study the historical relationships between languages and trace their evolutionary paths.

Genealogical Information in Selection Popup

User Story: As a linguistic professor, I want the selection popup to include more genealogical information, specifically the language family, so that I can quickly identify the relationship between languages.

Indicator of Polygons per Languoid

User Story: As a linguistic expert, I want the application to indicate the number of polygons per languoid on the map to understand the extent of its distribution.

Indicator of Existing Polygons in Tree Visualization

User Story: As an academic user, I need the tree visualization to include indicators showing whether polygons exist for each languoid, enabling me to distinguish between represented and non-represented languages.

Selectable Data Table for Map Entries

User Story: As a researcher, I want a selectable data table that showcases the entries on the map, allowing me to compare and analyze language data more efficiently.

Clearer and Intuitive View

User Story: As a linguistic professor, I need a clearer and more intuitive view of the application interface to make exploration and analysis easier.

Download Options for Language Data

User Story: As a user, I want download options for the language data either through a user interface or an API, facilitating further research and analysis.

Color-Coordinated Language Polygons for Tree Exploration

User Story: As a researcher, I require the language family trees to be color-coordinated to visually depict the relationships among languages, enhancing the understanding of language evolution.

Indigenous or Colonial Information

User Story: As a linguistic expert, I want information on whether a language is indigenous or colonial to contextualize its historical development and migration.

Polygons Without Borders for Clear View

User Story: As a user, I prefer polygons without borders on the map for a clearer view of language distribution and to avoid clutter.

Time slider for Temporal Filtering

User Story: As an user, I need a time slider to filter the polygons on the map for specific timeframes, aiding in the study of language changes over different periods.

Option to Pin Polygons for Comparison

User Story: As a linguistic researcher, I want the option to pin polygons on the map layer for easy comparison and analysis of different language distributions.

Hover Effect for Quick Language Information

User Story: As a user, I expect a hover effect on the polygons to display relevant language information, providing quick access to details.

Base Map with Relief for Geographic Context

User Story: As a user, I would like a base map with relief to have a better understanding of the geographic context while exploring language distributions.

Participant 2

Goals and Objectives:

- To study language variation and change in Spanish and Ibero-Romance languages through empirical and quantitative research.
- To explore morphosyntactic phenomena attested in social networks, such as Twitter, and analyze their impact on language evolution.
- To use language corpora, fieldwork data, and statistical tools to gather evidence and draw meaningful conclusions in her research.
- To find an application that provides detailed speaker count information and presents language data in a better version of the Ethnologue.

Needs and Preferences:

- Access to a comprehensive and well-curated language database with detailed speaker count information.
- A user-friendly interface that allows easy exploration of language variation and change on the map.
- Features to support quantitative analysis, statistical tools, and data visualization for research purposes.
- Reliable and up-to-date language data from various Ibero-Romance languages, especially Spanish.
- Customizable data filters to focus on specific linguistic phenomena and research interests

Challenges:

- Dealing with large and diverse language datasets, ensuring their accuracy and relevance for dialectological research.
- Identifying and analyzing morphosyntactic variation and change patterns across social networks like Twitter.
- Finding an application that presents language data in a way that is more useful and comprehensive than existing resources like the Ethnologue.

Scenario:

Participant 2 is engaged in a research project centered on morphosyntactic variation in Spanish and other Ibero-Romance languages. The project's aim is to analyze variation within social networks, necessitating a language visualization application equipped with reliable speaker count data and customizable data filters. Upon using the application, the participant finds it particularly beneficial for their research, thanks to its detailed speaker count information, user-friendly interface, and the capability to filter data according to specific linguistic phenomena. The application provides valuable insights and effectively supports their empirical and quantitative research methodology.

Key Takeaways:

Participant 2 is a dialectologist specializing in language variation and change, particularly in Spanish and other Ibero-Romance languages. They require a language visualization application that aligns with their quantitative research methods and offers detailed speaker count information. The application should also provide customizable data filters and user-friendly tools to facilitate efficient exploration and analysis of linguistic phenomena.

User Stories

Speaker Count in Information Tab

User story: As a dialectologist, I want the speaker count to be included in the information tab for each language, as it is of high relevance to my research.

Improved Language View for Easy Comparison

User story: As a researcher, I find the current language view unintuitive since the polygons disappear with every click or move, making it difficult to compare languoids. I want a better version that allows smooth exploration and comparison.

Hover Effect for Polygon Information

User story: As a user, I want a hover effect on the polygons, displaying relevant information about each language, so that I can quickly access details without clicking.

Keep Borders for Distinguishing Polygons

User story: As a dialectologist, I prefer the language polygons to have borders since they make it easier to distinguish between different language territories on the map.

Improved Source Selection and Comparison

User story: As a researcher, I desire a more intuitive way to select and compare sources, enabling efficient exploration and analysis of linguistic data.

Enhanced Exploration of Language Trees

User story: As a dialectologist, I would like more options to explore language trees, including not only direct ancestors but also the option to view sibling languages for a comprehensive analysis.

Selecting Languages and Their Predecessors

User story: As a user, I want the possibility to select a language and its predecessors to showcase on the map simultaneously, facilitating visual comparisons.

Beneficial Images for Source Understanding

User story: As a researcher, I am fond of the source view, and I believe including images of the sources would be beneficial for understanding the quality and context of the linguistic data.

Option for Viewport Pan to Showcased Polygon

User story: As a dialectologist, I find it interesting to have an option to pan the viewport of the map automatically to the showcased polygon, improving navigation and focus.

Short Descriptions for Source and Tree Dropdowns

User story: As a researcher, I suggest adding short descriptions for the source and tree dropdowns to provide better understanding and ease of navigation in the application.

Indicator for Presence of Language Polygon

User Story: As a dialectologist, I need the language visualization application to provide an indicator showing whether there is a language polygon available for a languoid on the map. This indicator will help me quickly identify and distinguish between represented and non-represented languages during my research and analysis.

Participant 3

Goals and Objectives:

- To deepen his understanding of language variation and change, especially in dialects and sociolinguistic contexts.
- To explore linguistic phenomena in diverse linguistic communities and uncover patterns of variation and change.
- To use data analysis and statistical tools to conduct empirical research in linguistics and contribute to the field's knowledge.

Needs and Preferences:

- Access to a reliable and comprehensive source of linguistic data to support his research and analysis.
- Tools and features that facilitate the exploration of language variation, including dialects and sociolinguistic factors.
- A user-friendly interface that enables efficient data manipulation, visualization, and statistical analysis.
- The ability to customize data views and filters to focus on specific linguistic phenomena and research questions.
- Reliable and up-to-date information on language speaker counts and distribution.

Challenges:

- Managing and analyzing large datasets of linguistic data while maintaining data quality and accuracy.
- Identifying and interpreting patterns of language variation and change in different linguistic communities.
- Finding accessible and user-friendly linguistic resources and tools to support his research.

Scenario:

A graduate student in linguistics, with a keen interest in sociolinguistics and dialectology, is currently engaged in a research project that investigates language variation within a specific dialect region. This student utilizes a language variation and change application to access linguistic data, explore dialectal differences, and analyze language variation. The application's user-friendly interface and customization options enable efficient filtering and visualization of data relevant to the research questions. The student also relies on the speaker count information provided by the application to gain a deeper understanding of the linguistic landscape of the study area.

Key Takeaways:

The graduate student, specializing in sociolinguistics and dialectology, places high value on accessible and user-friendly tools that facilitate the exploration of language variation and change, particularly in dialects. Essential resources for their research include a reliable source of linguistic data, customization options, and statistical tools. Access to detailed speaker count information is particularly appreciated, as it enhances understanding of the linguistic communities under study.

User Stories:

Improved Loading Screen in Info Tab

User Story: As a user, I find the loading screen in the info tab confusing. I want a clearer and more informative loading screen that provides feedback on the loading progress.

Enhanced Initial Experience with Polygon Display

User Story: As a user, I find it unintuitive when no polygons are shown at the start. I expect to see at least some initial information or visual representation when I open the application.

Improved Tree View Indicators

User Story: As a user, I'm not sure what is being shown in the tree view. I would like better indicators or labels to understand the context of the tree view.

Focus on Specific Levels of Languoids

User Story: As a researcher, I typically focus on specific levels of languoids rather than language families. I want the option to easily navigate and explore specific levels of the language hierarchy.

Enhanced Comment Field with Source Information

User Story: As a user, I want a comment field that includes a picture of the source and minimal information about what it represents, making it easier to understand the source context.

Fix Errors in Switching Between Map Views

User Story: As a user, I encounter errors when switching between map views. I want a smoother and error-free experience when transitioning between different map views.

Separation of Language Families and Languages on the Map

User Story: As a user, I find it problematic when language families and individual languages are mixed on the same map. I prefer a clear separation or differentiation between them.

Exclude Bookkeeping Elements from the Prototype

User Story: As a user, I want bookkeeping elements to be excluded from the prototype to ensure a cleaner and more focused user experience.

Dynamic and Visible Legend

User Story: As a user, I expect the legend to be dynamic and clearly visible, providing information about the map's visual elements.

Improved Navigation Options

User Story: As a user, I find navigation difficult because I can only gather information by clicking on the map. I want additional navigation options to access language information more easily.

Clear Clickability of Sources

User Story: As a user, I find it unclear that sources are clickable. I want a clear indication that sources can be clicked for more information.

Clarify Urheimat Representation

User Story: As a user, I need clarification on the representation of "Urheimat." I want to know if it refers to the total extent or the historical part where the protolanguage was spoken.

Enhanced Tree Filtering Options

User Story: As a user, I want more filtering options for the tree view, including the ability to filter by branches, siblings, lineage, and other criteria.

Incorporate Temporal Information

User Story: As a user, I want access to temporal information to understand the historical context of language data.

Display Information on Map Startup

User Story: As a user, I find a blank default map not intuitive. I prefer to see some initial information or visual representation when I start the application.

Options for Non-Overlapping Groups and Language Extent

User Story: As a user, I want options for handling non-overlapping groups of languages and differentiating opinions about a language's extent on the map.

Summarize and Simplify Clickable Polygons

User Story: As a user, I prefer not to have individual polygons for each source clickable. Instead, I want them summarized and simplified for easier interaction.

Enable Selection of Languages via Fuzzy Search

User Story: As a user, I want the ability to select languages via fuzzy search, making it easier to find and explore specific languages of interest.

Participant 4

Goals and Objectives:

- To advance her doctoral project by applying geospatial methods to conduct integrative and multidisciplinary research.
- To contribute to the understanding of the history of Finno-Ugric-speaking peoples and conduct various analyses related to northern Siberia.
- To develop and maintain the URHIA platform as a comprehensive repository for spatial data, enabling storage, sharing, and visualization of research findings.
- To ensure effective communication between scientists and IT development teams to enhance the platform's usability and meet the needs of its users.

Needs and Preferences:

- Access to geospatial tools and methods for conducting complex integrative research.
- A platform that facilitates the storage, sharing, and visualization of spatial data on a map.
- Effective communication and collaboration with interdisciplinary teams, including linguists, archaeologists, geneticists, and IT developers.
- A user-friendly and intuitive platform that supports research in the history of Finno-Ugric-speaking peoples and analyses of northern Siberia.

Challenges:

- Managing and analyzing diverse and multidisciplinary datasets to draw meaningful insights.
- Ensuring effective communication and collaboration between researchers with different backgrounds and expertise.
- Developing and maintaining a user-friendly spatial data platform that meets the evolving needs of scientists.

Scenario:

A doctoral student is currently advancing their project, which encompasses geospatial analysis and integrative research on the history of Finno-Ugric-speaking peoples and northern Siberia. This student collaborates actively with experts from linguistics, archaeology, genetics, and IT to maintain and enhance the URHIA platform, a spatial data platform they co-developed. This platform serves as a crucial resource for storing, sharing, and visualizing spatial data pertinent to their research areas. Through engaging in effective dialogue with interdisciplinary teams, the student ensures that the platform’s functionalities are well-aligned with the diverse needs of its users, thus contributing to the success of various research projects.

Key Takeaways:

The doctoral student is an interdisciplinary researcher with expertise in geospatial analysis and integrative research. A significant accomplishment includes the collaborative development of the URHIA platform, which facilitates the storage, sharing, and visualization of spatial data for research in linguistics, archaeology, genetics, and other disciplines. Effective communication and collaboration are key to their work, ensuring the platform continuously meets the evolving needs of its diverse user base.

User Stories:

Image/file of Source is Desirable

User Stories: As a user, I want the ability to access and download an image or file of the data source, so that I can have a tangible reference for the data's origin and authenticity.

Intuitive Access to Attribute Table

User Stories: As a user, I want to easily access and filter the attribute table, so I can quickly find the data I need without going through multiple steps.

Fly-To Functionality

User Stories: As a user, I need a 'Zoom to Selection' feature, so that I can focus on and analyze specific areas or features on the map more effectively.

Facilitation of Maintaining the Visibility of Features when Clicking on Different Feature

User Stories: As an analyst, I want the map to maintain the visibility of selected elements when clicking elsewhere, so I can compare different data points without losing my current selection.

Inclusion of Data Table in the Map View

User Stories: As a user, I want a seamless integration of the data table with the map view, so that I can filter for new map features without change the map view.

API for Programmatic Data Access

User Stories: As a data analyst, I need direct connectivity to databases from the application, so I can query and retrieve data efficiently without unnecessary downloads.

Search by Location

User Stories: As a user, I want to search and zoom into specific geographical locations easily, to quickly find the area I am interested in.

Real-time Data Filtering (in search process)

User Stories: As a researcher, I need an interactive and real-time data filtering system on the map, so that I can dynamically explore the data based on my search criteria.

Add Options for Base Map

User Stories: As a User, I want access to different base maps in order to choose the best way to view and present the data according to my needs.

Tutorials for New Users

User Stories: As a new user, I want guidance and tutorials, so that I can learn how to effectively use the application for complex analyses or visualizations.

Appendix C – Task-Based Scenarios

Tutorials and Descriptive Dropdowns

Provides initial guidance for new users through an informational panel and clear, informative dropdown guide to enhance starting experiences.

Task: "I'd like you to get acquainted with the map by finding information on a language you're interested in. How would you go about starting this task?"

- **Display Information on Map Startup:** Present essential information or tips when the map initially loads to orient users.
- **Short Descriptions for Source and Tree Dropdowns:** Provide concise descriptions within dropdown menus to assist user selection.
- **Tutorials for New Users:** Offer introductory tutorials that guide new users through the platform's features and capabilities.

Checklist:

- ✓ Participant browses initial map information and explains or demonstrates how they would initiate a search for specific language information.
- ✓ Participants reads map guide for orientation.
- ✓ Participant then proceeds to browse the language information component.

Map Relief and Base Map Options

Directly addresses improving the map's visual context for better geographic understanding and personalization.

Task: "Picture you're highlighting languages in diverse environments. See if adjusting the map's appearance aids in distinguishing these areas."

- **Base Map with Relief for Geographic Context:** Adjust map background to include geographic relief, aiding spatial understanding.
- **Add Options for Base Map:** Provide users with the ability to customize their map view by selecting different base maps.

Checklist:

- ✓ Participant locates and uses the map background options.
- ✓ Participant selects and applies a different base map.

Quick Information on Hover

Enables immediate insight into language and geographical details through hovering, streamlining user interaction for efficiency.

Task: "As you browse the map, let's say you come across a region dense with diverse languages. Explore the map to find a quick snapshot of linguistic data for areas that pique your interest."

- **Hover Effect for Quick Language Information:** Display language names or brief info upon hovering over a language area.
- **Hover Effect for Polygon Information:** Show specific information about the language polygon (e.g., name, dialect) on hover.
- **Keep Borders for Distinguishing Polygons:** Maintain clear borders around language areas for better distinction and clarity.

Checklist:

✓ Participant hovers over a language area to trigger the quick info display.
✓ Participant identifies the type of language or dialect from the hover info.
✓ Participant differentiates polygons using the borders maintained for clarity.

Visual Sources for Context

Integrates images and source documentation with the map, providing contextual background to support language data.

Task: "Consider you're verifying the historical context of the language data presented. How would you go about finding visual aids or source materials that could reveal this information?"

- **Beneficial Images for Source Understanding:** Integrate images that contextualize or support the language data presented.
- **Image/file of Source is Desirable:** Enable access to original source files or images directly from the map interface for deeper exploration.

Checklist:

✓ Participant switches to the Map view in the information component.
✓ Participant finds the source image and additional source information.
✓ Participant selects a image to have a larger display of source map image.

Search and Real-time Filtering

Combines tools for searching languages and filtering results on the fly, making language discovery more accessible.

Task: "Let's say you've just learned about the 'Venezuelan-Cariban' language family and you're eager to delve deeper into its languages. Find a way on the map to narrow down your search to this particular family."

- **Enable Selection of Languages via Fuzzy Search:** Implement a search feature that allows for the approximation in language names, facilitating discovery.
- **Real-time Data Filtering (in search process):** Offer filters during search to narrow down results dynamically based on user criteria for languages and sources.

Checklist:

- ✓ Find search and open it.
- ✓ Participant enters a search term related to the 'Venezuelan-Cariban' language family in the search bar.
- ✓ Participant selects the 'Venezuelan-Cariban' family from the search suggestions.
- ✓ Review the responsiveness of the system to real-time filtering and the relevance of the filtered data.

Language Trees and Relationships

Offers users tools to navigate through language families, understand relationships, and delve into linguistic history.

Task: "Imagine you've encountered a new language and you're curious about its lineage and connection to other languages. Investigate how you might uncover these genealogical links using the map."

- **Enhanced Exploration of Language Trees:** Provide interactive tools to explore the connections within language families.
- **Analyze Language Relationships and Family Trees:** Enable detailed examination of how languages relate within their families.
- **Selecting Languages and Their Predecessors:** Allow users to trace and select languages along with their historical lineage.
- **Focus on Specific Levels of Languoids:** Permit users to drill down into specific levels of language detail (e.g., from family to dialect).
- **Indicator of Existing Polygons in Tree Visualization:** Highlight which languages in the tree have corresponding polygons on the map.
- **Indicator for Presence of Language Polygon:** Pictogram when a selected language has a visible representation on the map.

- **Improved Tree View Indicators:** Enhance visual cues within the language tree to guide user exploration effectively.

Checklist:

- ✓ Participant locates language family tree.
- ✓ Participant identifies the selected language as the highlighted language node in the tree.
- ✓ Participant traces and selects languages and their historical predecessors in the tree.
- ✓ Participant observes and utilizes the indicator in next to the languoid name.
- ✓ Participant understands what interacting with the indicator results in.

Visibility, Comparison, and Legends

Focuses on maintaining clarity while interacting with multiple languages and improving comparative analysis tools.

Task: "You're analyzing linguistic data for a report and need to compare several languages. Determine how you might keep track of and compare your findings within the map."

- **Facilitation of Maintaining the Visibility of Features when Clicking on Different Features:** Ensure that selected features remain visible and distinct even when exploring new selections.
- **Option to Pin Polygons for Comparison:** Allow users to "pin" selected language areas for side-by-side comparison.
- **Improved Language View for Easy Comparison:** Optimize the interface for comparing multiple languages or dialects simultaneously.
- **Dynamic and Visible Legend:** Implement a legend that updates dynamically based on the map view, aiding interpretation.

Checklist:

- ✓ Participant pins language polygons to compare linguistic data through the legend.
- ✓ Participant pins language polygons to compare linguistic data through the language information panel.
- ✓ Participant reads and interprets the dynamic legend to understand what map features are displayed.
- ✓ Participant understands the highlighted languages represent pinned language polygons.

Source Selection Clarity

Aims to make the process of choosing and understanding data sources more transparent and user-friendly.

Task: "Assume you're researching the linguistic landscape of the Amazon basin and know multiple scholars have mapped it differently. Locate these varied source maps on the platform and assess how each depicts the area."

- **Improved Source Selection and Comparison:** Simplify the process for selecting and comparing different language data sources.
- **Clear Clickability of Sources:** Make it evident which elements in the interface are interactive for accessing source information.

Checklist:

✓ Participant locates source dropdown in the language view.
✓ Participant selects different data sources and updates language polygon.
✓ Participant compares information from multiple sources for clarity.

Navigation and Viewport Adjustments

Enhances user control over map navigation and the ability to adjust perspectives easily.

Task: "You're investigating the historical migration of the Tupi-Guarani language family but can't find where they're located on the map. Discover how to employ a specific feature, like 'Fly-To,' to bring you directly to the relevant area or polygons representing this language family."

- **Fly-To Functionality:** Implement a feature that smoothly transitions the map view to a selected language area.
- **Option for Viewport Pan to Showcased Polygon:** Enable easy panning to areas of interest, especially after a search or selection.

Checklist:

✓ Participant locates Fly-To functionality in the legend.

✓ Participant successfully uses 'Fly-To' functionality in the legend to navigate to a language area.

✓ Participant discovers that through updating language or source selection either through the fuzzy search and language tree adjust the viewport to the language polygon.

Locating Language Data for a Specific Location

Task Instructions for Participant: "Please show me how you would find out what language data we have available for the town of Ivirgarzama in Bolivia using the map's features."

- **Checklist**

- Participant utilizes the search function or map navigation to locate Ivirgarzama, Bolivia.
- Participant finds and describes the language data available for Ivirgarzama, Bolivia.
- Participant uses Ctrl + Cursor Left Click to locate what languages are spoken in the area where the participant clicked on.

- **Requirement Covered:**

  Search for location: Enable the location search for easy navigation in lesser known regions.

Reliable Access and Data Download

Focuses on system reliability improvements and introduces functionalities for downloading data and accessing APIs.

- **Fix Errors in Switching Between Map Views:** Address and correct any issues encountered when changing between different map presentations.
- **Improved Loading Screen in Info Tab:** Optimize the loading interface for accessing detailed language information, enhancing user patience and experience.
- **Download Options for Language Data:** Provide functionality for users to download language data for offline use or further analysis.
- **API for Programmatic Data Access:** Offer an API for developers or researchers to access language data programmatically, facilitating external analysis and integration.

Checklist:

✓ Participant encounters and resolves any navigation issues between map views.

✓ Participant engages with the info tab and evaluates loading screen efficiency.

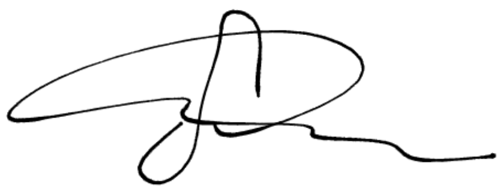
✓ Participant downloads language data and acknowledges the download process.

✓ Participant explores the API for programmatic data access and indicates understanding.

Personal declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Zurich, 30 April 2024



Silvan Caduff