ALGORITHMIC STANDARDIZATION OF TOPONYMS IN NAVIGATION APPLICATIONS: A MULTI-LEVEL ANALYSIS OF THE INTERACTION BETWEEN USERS, PLATFORMS AND SYMBOLIC SPACE—PART I

Carmen-Maria ȚEPEȘ
c.m.tepes@gmail.com

Abstract: *This paper analyses the dynamics of place names in the context of the digitization of orientation practices, focusing on the relationship between linguistic usage, standardization processes, and the constraints imposed by digital platforms. Within an interdisciplinary theoretical framework, the research approaches toponymy as a space of intersection between norm and variation, highlighting the role of endonyms and exonyms, of official and vernacular forms, as well as of diatopic and diastratic variation. It also discusses mechanisms of linguistic standardization and their implications in relation to language as a social practice, anchored in power relations and spatial representation. Methodologically, the study employs a mixed-methods approach, combining qualitative analysis of interviews with quantitative analysis of a corpus of toponyms extracted from digital applications. Patterns of lexical, graphic, and pragmatic variation are investigated, as well as processes of standardization and variation reduction. Multilevel modelling allows for the identification of individual and contextual influences on the use of toponyms. The results indicate a general preference for standardized forms, though vernacular forms persist alongside them. Digital platforms play an essential role in the selection and ranking of toponymic forms, favouring institutional names and contributing to linguistic standardization. At the same time, these processes generate tensions between everyday usage and algorithmically imposed norms. In conclusion, this paper highlights the role of digital platforms as active agents of linguistic standardization, emphasizing the need for a critical understanding of how they reconfigure toponymic practices and the relationship between language and space.*

Keywords: *toponymy, interdisciplinarity, linguistics, geography, cultural identity, collective memory, etymology, digitalization, statistical analysis.*

INTRODUCTION

1.1. The digitalization of wayfinding practices

In recent years, everyday spatial orientation relies increasingly on digital platforms, which gradually displace previous forms of spatial knowledge, grounded in direct experience

and locally transmitted practices (Farman, 2012; Dodge & Kitchin, 2007; Kitchin et al., 2017; Leszczynski, 2019; Wilmott, 2020). Applications such as Google Maps and Waze have become the main interfaces through which users access, organize, and interpret space-related information. This shift extends beyond the adoption of new tools and it reflects a broader reconfiguration of how space is known and represented. Spatial knowledge is increasingly accessed through pre-structured, platform-mediated representations that prioritize consistency, legibility, and immediate usability (Farman, 2012; Leszczynski, 2019; Kitchin et al., 2017; Dodge & Kitchin, 2007). In this sense, digital mapping environments are involved in the production of spatial knowledge by framing what is visible, searchable, and linguistically available (Graham & Zook, 2013; Zook & Graham, 2007).

From a linguistic perspective, this transformation affects how toponyms are acquired, transmitted and stabilized. Traditionally, place names emerge within communities of practice and are transmitted through everyday interaction, where they reflect shared histories, habitual usage, and locally embedded knowledge. Such forms are often variable and context-dependent, and their use may index familiarity, social belonging, or specific patterns of spatial experience (Tent & Blair, 2011; Alderman, 2008; Ainiala et al., 2012). Toponymic variation, including the coexistence of official and vernacular forms, is therefore a normal feature of linguistic practice rather than a deviation from it (Berg & Vuolteenaho, 2009; Rose-Redwood et al., 2010).

Digital mapping platforms, however, operate on principles that prioritize stability and standardization. Place names are stored as discrete entries in structured databases, where each location is typically associated with a single preferred form. Users encounter these forms repeatedly in uniform contexts—search results, navigation instructions, and map labels—which limits exposure to alternative variants and reinforces a narrow range of acceptable forms (Graham et al., 2014; Shelton et al., 2015). As a result, the circulation of toponyms becomes increasingly modelled by platform logics, including their frequency of use, algorithmic ranking, and commercial visibility (Light & Young, 2015; Wilken & Goggin, 2015).

This shift has direct implications for linguistic usage. Instead of acquiring place names primarily through interactions within a community, users increasingly adopt the forms they come across through digital interfaces. Repetition and prominence give legitimacy, and the most visible forms become the most readily used in everyday discourse. In this way, digital platforms act not only as repositories of spatial data, but also as mechanisms of linguistic selection, contributing to the consolidation of certain toponymic variants while marginalizing others (Rose-Redwood et al., 2010; Berg & Vuolteenaho, 2009; Graham & Zook, 2013). In time, this process may lead to a reduction in toponymic variation and to a weakening of the indexical dimension of place names, as locally meaningful forms become less accessible within dominant channels of representation (Ainiala et al., 2012; Leszczynski, 2019).

1.2. The standardization of toponyms in algorithmic environments

The use of toponyms in large-scale digital systems depends on a high degree of consistency, stability, and formal clarity. To function effectively, such systems require each spatial entity to be associated with a clearly identifiable and unambiguous linguistic label. This requirement leads to processes of normalization and disambiguation, through which multiple coexisting variants are aligned and reduced to a single preferred form. Where

variation exists—whether in the form of alternative spellings, vernacular denominations, or context-dependent usages—it is typically resolved through selection rather than preservation. While these processes are necessary for database structure and search efficiency, they also limit the range of forms that remain visible and usable within the system (Dodge & Kitchin, 2007; Kitchin et al., 2017; Graham et al., 2014).

From a linguistic perspective, this shift has direct implications for the treatment of toponymic variation. In non-digital contexts, the coexistence of multiple toponyms for the same location is both common and functionally meaningful. Different forms may reflect degrees of familiarity, social positioning, or patterns of local usage, and their distribution is often sensitive to context. By contrast, digital systems tend to suppress such variation, favouring a single standardized form that can circulate across contexts without ambiguity. In doing so, they reframe variation as a problem to be resolved rather than a feature of linguistic practice (Ainiala et al., 2012; Tent & Blair, 2011; Berg & Vuolteenaho, 2009).

Within this framework, standardization can no longer be attributed exclusively to institutional actors such as state authorities or cartographic bodies. It increasingly follows the internal logic of digital platforms, where the prominence of a given toponym depends on factors such as frequency of use, user-generated input, and commercial relevance. Toponyms that are more frequently searched, more widely used, or more closely tied to commercial entities tend to gain greater visibility, while less common or locally specific variants are less likely to appear (Graham & Zook, 2013; Shelton et al., 2015; Graham et al., 2014; Zook & Graham, 2007).

As a result, digital platforms take on a normative role in shaping toponymic usage. Through repeated exposure to search results, map displays, and route instructions, certain forms become established as the default options available to users. Over time, this repeated exposure contributes to their stabilization in everyday language. At the same time, toponyms that do not align with platform criteria—particularly vernacular or context-dependent forms—become less accessible and may gradually fall out of use. In this way, the constraints of digital representation extend beyond technical considerations, influencing not only how toponyms are stored and displayed, but also how they circulate and are reproduced in everyday discourse (Light & Young, 2015; Leszczynski, 2019; Wilmott, 2020; Ainiala et al., 2012).

1.3. Research objectives and questions

This study investigates the role of digital mapping platforms in the selection, stabilization, and circulation of toponyms, with particular attention to their effects on patterns of linguistic variation and use. The analysis is grounded in the assumption that platform-mediated environments do not simply display toponyms, but actively structure their distribution, visibility, and uptake in everyday discourse.

The primary objective of the study is to identify the mechanisms through which digital platforms privilege certain toponymic forms over others and to explain how these mechanisms contribute to the stabilization of specific variants. This involves examining the interplay between database constraints, algorithmic ranking, and patterns of user interaction, and assessing how these factors shape the relative visibility of competing forms.

A second objective is to analyse the relationship between exposure and usage, focusing on the extent to which repeated encounters with standardized toponyms in digital interfaces lead to their adoption in active linguistic practice. The study seeks to determine

whether platform-promoted forms remain confined to recognition and passive use, or whether they become integrated into speakers' productive repertoire, potentially displacing alternative variants.

A third objective is to assess the impact of digital standardization on vernacular toponyms and on forms with strong indexical value. The analysis considers how such forms are affected by reduced visibility within platform environments, and whether their marginalization in digital contexts translates into decreased circulation in everyday interaction.

A further objective is to account for variation across users and contexts. The study examines how factors such as intensity of platform use, degree of local familiarity, and characteristics of the spatial setting condition the adoption and use of standardized toponyms, and whether different user profiles exhibit distinct patterns of alignment with platform-preferred forms.

In line with these objectives, the research is structured around four interconnected lines of inquiry. First, it seeks to establish how digital platforms organize the selection and stabilization of toponyms, by identifying the criteria and processes through which certain forms are made more visible, and others are excluded or deprioritized. Second, it examines the relationship between exposure and linguistic behaviour, with a view to determining whether repeated contact with standardized forms leads to measurable changes in usage. Third, it analyses the position of vernacular and context-dependent toponyms within this system, focusing on the conditions under which they persist, are reconfigured, or gradually disappear. Finally, it investigates the extent to which these processes vary across user groups and spatial contexts, to capture the differential effects of platform mediation on toponymic practice.

1.4. Conceptual delimitations

Toponyms are understood here as linguistic forms that fulfil both referential and indexical functions, anchoring spatial reference, while simultaneously signalling forms of social knowledge, familiarity, and situated use. They do not operate as neutral labels, but as elements embedded in communicative practices, reflecting shared experience and patterns of interaction within specific communities. In this sense, toponymy is approached as a mutable system, shaped by geographic distribution, social differentiation, and pragmatic context, rather than as a fixed inventory of rigid forms (Tent & Blair, 2011; Ainiala et al., 2012; Berg & Vuolteenaho, 2009).

Standardization is defined as the process through which some toponymic forms are selected, stabilized, and reproduced within a field of competing variants. While traditionally associated with institutional regulation and cartographic practices, standardization is extended in this study to include processes driven by digital infrastructures. In this context, selection is not only the result of explicit norm-setting, but also of patterned visibility, repetition, and use within platform environments (Milroy & Milroy, 2012; Graham & Zook, 2013).

Algorithmic environments are understood as digital systems in which linguistic forms are stored, organized, and then displayed according to computational procedures and operational constraints. Within such systems, toponyms are treated as discrete entries that must meet some requirements of consistency, uniqueness, and retrievability. Their circulation is shaped by mechanisms such as ranking, filtering, and user interaction, which determine which forms become more visible and, therefore, more likely to be adopted. These environments thus influence not only how toponyms are represented, but also how

they are recognized and used in everyday discourse (Kitchin et al., 2017; Leszczynski, 2019; Graham et al., 2014).

2. THEORETICAL FRAMEWORK

2.1. Toponymy, between usage, norm, and variation

Toponyms stand at the intersection of linguistic structure and social practice, combining relatively stable formal properties with a high degree of variability in use. Although they are often seen as fixed referential labels within administrative and cartographic frameworks, their actual functioning in language is shaped by ongoing interaction between institutional norms and everyday communicative practices. Official naming systems establish standardized forms through processes of codification and regulation, yet these forms coexist with a range of alternative variants that circulate in informal contexts and reflect locally grounded patterns of usage (Berg & Vuolteenaho, 2009; Rose-Redwood et al., 2010; Alderman, 2008).

From a linguistic perspective, toponyms are not merely designators of spatial entities, but elements embedded in systems of variation. Their use is sensitive to factors such as speaker position, degree of familiarity with a place, and communicative context. Different forms may be selected depending on whether the speaker addresses an insider or an outsider audience, engages in formal or informal interaction, or draws on locally shared knowledge. In this sense, variation in toponyms is not incidental, but functionally meaningful, contributing to the expression of social relations and patterns of spatial experience (Ainiala et al., 2012; Tent & Blair, 2011; Hough, 2016).

At the same time, the distribution of toponymic forms reflects broader processes of norm formation and negotiation. Institutional authorities, including state agencies and cartographic bodies, play a central role in defining official forms, which are then disseminated through education, administration, and formal mapping practices. However, these forms do not necessarily displace vernacular usage. Instead, they enter into a dynamic relationship with locally used variants, which may persist, adapt, or coexist alongside official denominations. This coexistence highlights the fact that toponymy operates as an open and evolving system, rather than a closed set of fixed labels (Berg & Vuolteenaho, 2009; Rose-Redwood et al., 2010; Light & Young, 2015).

In this light, toponyms can be understood as sites where linguistic, social, and spatial processes converge. Their apparent stability at the level of form often masks a more complex reality, in which meaning, usage, and distribution are continuously renegotiated. The tension between standardization and variation is therefore inherent to toponymic systems, reflecting the broader interplay between normative regulation and everyday linguistic practice (Alderman, 2008; Ainiala et al., 2012; Hough, 2016).

2.1.1. Endonyms and exonyms

The distinction between endonyms and exonyms captures two fundamentally different perspectives on the naming of places and the circulation of toponyms across linguistic and cultural boundaries. Endonyms refer to the forms used by a community to designate places within its own linguistic and social space, and are therefore closely tied to local usage, historical continuity, and shared cultural knowledge. Exonyms, by contrast, are forms used outside that community, typically emerging through processes of translation,

phonological adaptation, or historical convention, and reflecting external perspectives on the same referential entities (Jordan, 2012; Kadmon, 2000).

From a linguistic standpoint, the relationship between endonyms and exonyms is not merely one of formal difference, but also of functional and pragmatic variation. Endonyms tend to carry stronger indexical value, signalling familiarity with a place and participation in a given linguistic community, while exonyms often facilitate communication across linguistic boundaries, prioritising accessibility and recognisability over local specificity. The coexistence of these forms illustrates the inherently plural character of toponymic systems, in which multiple naming practices can operate simultaneously, depending on context and audience (Tent & Blair, 2011; Ainiala et al., 2012).

In contemporary digital environments, however, this balance is increasingly affected by platform-driven processes of selection and display. Digital mapping systems tend to privilege forms that are widely recognisable, standardised, and easily retrievable, often aligning with either official endonyms or internationally established exonyms. The criteria governing this selection are not purely linguistic, but are shaped by factors such as frequency of use, global visibility, and user search behaviour. As a result, locally embedded forms may be overshadowed by variants that circulate more widely within platform ecosystems, regardless of their status in everyday usage (Graham & Zook, 2013; Leszczynski, 2019; Wilmott, 2020).

This shift has major implications for the distribution and stability of endonyms and exonyms. By reinforcing certain forms through repeated display and patterned visibility, digital platforms contribute to the consolidation of specific variants, while others become less accessible and less frequently used. The ranking and filtering mechanisms that structure access to toponyms favour forms that are more widely used, more easily retrievable, or more closely aligned with dominant linguistic and economic patterns. As a result, visibility becomes a key factor in the stabilization of toponymic forms, with direct consequences for their circulation in everyday language (Graham & Zook, 2013; Graham et al., 2014; Shelton et al., 2015; Leszczynski, 2019; Wilmott, 2020).

2.1.2. Official and vernacular toponyms

Official toponyms are established through formal administrative procedures and are codified in legal documents, official registers, and cartographic representations. Their form and use are regulated by institutional frameworks, which aim to ensure consistency, standardization, and broad recognisability across contexts. As such, official toponyms tend to circulate in domains associated with governance, planning, and formal communication, where stability and uniformity are required (Alderman, 2008; Berg & Vuolteenaho, 2009).

Vernacular toponyms, by contrast, arise within everyday speech and are shaped by local patterns of interaction and shared spatial experience. They often reflect informal naming practices, historical references, or socially salient landmarks, and may vary across speaker groups or communicative situations. These forms frequently carry strong indexical meanings, signalling familiarity with a place and membership in a local community. Their distribution is therefore more fluid and context-sensitive, and their use may depend on factors such as audience, setting, or degree of local knowledge (Tent & Blair, 2011; Ainiala et al., 2012; Hough, 2016).

The relationship between official and vernacular toponyms is not one of simple replacement, but of coexistence and functional differentiation. However, vernacular forms

are generally less stable at the level of form and less widely codified, which makes them more difficult to incorporate into systems that require fixed, unambiguous labels. In digital environments, where toponyms must be stored as discrete and standardised entries, this lack of stability and formalisation poses practical constraints. As a result, vernacular toponyms are less likely to be represented, or are incorporated only selectively, often losing part of their contextual specificity in the process (Graham et al., 2014; Leszczynski, 2019; Light & Young, 2015).

2.1.3. Diatopic and diastratic variation

Variation in toponymy reflects both geographic distribution and social differentiation, and constitutes an integral aspect of how place is linguistically organised and experienced. Different groups may employ distinct toponyms for the same location, depending on factors such as degree of familiarity, patterns of mobility, social positioning, or linguistic background. These differences are not random, but follow identifiable patterns linked to community-specific practices, local histories, and modes of spatial interaction. As such, toponymic variation can be understood as a structured phenomenon, shaped by the interplay of diatopic and diastratic factors (Ainiala et al., 2012; Tent & Blair, 2011; Hough, 2016).

From a pragmatic perspective, the choice of a particular toponym is often sensitive to context, including the relationship between speakers and the communicative setting. Speakers may shift between forms in order to accommodate different audiences, signal local knowledge, or ensure mutual intelligibility. In this sense, variation in toponyms performs important communicative functions, allowing speakers to position themselves socially and spatially within interaction (Alderman, 2008; Berg & Vuolteenaho, 2009).

However, such variation is frequently reduced or suppressed in standardized representations, particularly within institutional and digital contexts. Systems that require stable and unambiguous identifiers tend to privilege a single form, thereby masking the coexistence of alternative variants. This reduction is not merely a technical adjustment, but has consequences for the visibility and circulation of different toponymic forms, as those not selected are less likely to appear and, over time, less likely to be used (Graham et al., 2014; Leszczynski, 2019; Light & Young, 2015).

2.2. Linguistic standardization

2.2.1. Processes of norm formation and institutionalization

Classical approaches to linguistic standardization describe a set of interrelated processes through which particular forms are selected, codified, and disseminated within a speech community, typically under the authority of institutional actors (Milroy & Milroy, 2012; Haugen, 1966). These processes involve the identification of preferred variants, their formal stabilization in normative frameworks, and their diffusion through education, administration, and other channels of regulated use. Standardization is thus not only a linguistic process, but also a social and political one, closely tied to questions of authority, legitimacy, and control over language (Joseph, 1987; Milroy, 2001).

In the case of toponyms, these processes have traditionally been carried out by official bodies such as national mapping agencies, administrative authorities, or international standardization committees. Through formal procedures, these institutions establish the accepted forms of place names, which are then used in official documents,

maps, and public signage. Such forms are intended to ensure consistency and clarity across different domains of use, particularly in contexts where precise spatial reference is required (Kadmon, 2000; Berg & Vuolteenaho, 2009).

At the same time, the authority of institutional standardization has always coexisted with patterns of everyday usage, which may diverge from officially prescribed forms. This coexistence highlights the fact that standardization does not eliminate variation, but rather establishes a dominant norm within a broader field of competing variants. The accepted form gains legitimacy through institutional support, but its actual uptake depends on patterns of use within the speech community (Ainiala et al., 2012; Tent & Blair, 2011).

2.2.2. Reduction of variation and selection of dominant forms

Standardization involves the preferential selection of certain forms within a field of variation, resulting in the gradual reduction of alternative variants. This process does not simply eliminate variation, but reorganizes it by establishing a hierarchy in which some forms acquire normative status while others become peripheral. The selection of dominant variants is shaped by considerations of clarity, recognisability, and institutional legitimacy, and is typically reinforced through repeated use across formal domains (Milroy & Milroy, 2012; Joseph, 1987).

In digital environments, these dynamics are intensified by technical constraints that require unambiguous reference and efficient retrieval. The use of unique identifiers, the need for consistent database entries, and the optimization of search functions all favour the stabilization of a single preferred form for each referent. As a result, variation is not accommodated but managed through processes of alignment and reduction, whereby competing forms are either merged, excluded, or rendered secondary (Kitchin et al., 2017; Graham et al., 2014).

The outcome is the consolidation of a relatively limited set of dominant toponymic forms, whose status is further reinforced through algorithmic repetition and patterned visibility. Forms that are consistently displayed in search results, map interfaces, and route instructions gain additional stability by virtue of their frequency and accessibility. Over time, this repeated exposure contributes to their normalization in everyday usage, while less visible variants are progressively marginalised (Graham & Zook, 2013; Leszczynski, 2019; Shelton et al., 2015).

2.3. Language, space, and power

2.3.1. Toponyms as discursive practices

Toponyms are not neutral descriptors of space, but elements embedded in broader discursive processes through which social relations, historical trajectories, and symbolic hierarchies are articulated. The act of naming does not simply identify a location; it participates in the construction of meaning by selecting and foregrounding particular interpretations of place. As such, toponymic practices are closely tied to issues of representation, memory, and authority, reflecting the perspectives of those who have the power to name and to stabilize those names in public use (Berg & Vuolteenaho, 2009; Rose-Redwood et al., 2010; Azaryahu, 2011).

From this perspective, naming practices can both reinforce and contest existing structures of power. The institutionalization of specific toponyms—through official registers, maps, or public signage—often privileges certain historical narratives while

marginalizing others. Conversely, alternative or vernacular forms may function as sites of resistance, preserving locally meaningful references that are not captured by official designations. The coexistence of these forms highlights the inherently contested nature of toponymy, where different naming practices encode competing claims over space and meaning (Alderman, 2008; Light & Young, 2015).

In this sense, toponyms can be understood as part of a wider discursive field in which visibility is unevenly distributed. The forms that are most widely recognized and circulated are those that have been successfully institutionalized or otherwise rendered prominent, while others remain marginal or context-bound. The selection and stabilization of particular toponyms therefore have direct implications for which meanings and histories are made visible, and which are rendered less accessible within public and linguistic space (Berg & Vuolteenaho, 2009; Rose-Redwood et al., 2010; Azaryahu, 2011).

2.3.2. Indexicality and spatial anchoring

Beyond their referential function, toponyms carry a strong indexical dimension, in that they point to social context, local knowledge, and forms of belonging. Their use is not limited to identifying a location, but also signals degrees of familiarity, patterns of interaction, and participation in shared spatial practices. In this sense, the selection of a particular toponym can indicate whether a speaker is locally embedded, how they position themselves in relation to others, and what kind of knowledge they assume to be shared within a given interaction (Silverstein, 2003; Ochs, 1992; Ainiala et al., 2012).

Indexical meaning is closely tied to variation, as different forms may carry distinct social and pragmatic associations. Vernacular toponyms, for instance, often encode locally salient references or historically grounded meanings that are not transparent to outsiders, while more formal or standardized forms tend to operate across a wider range of contexts. The choice between such variants is therefore not neutral, but reflects communicative strategies and social positioning, as speakers adjust their usage in relation to audience, setting, and purpose (Alderman, 2008; Hough, 2016).

In digital environments, however, this indexical dimension is often reduced. Standardized forms are selected and displayed in ways that prioritize clarity, uniformity, and broad accessibility, which limits the visibility of context-dependent variants. As a result, toponyms are increasingly presented as stable, context-independent labels, detached from the social practices in which they are embedded. This shift does not eliminate indexical meaning altogether, but tends to weaken its role in shaping usage, as forms that carry strong local or interactional associations become less accessible within dominant channels of representation (Graham et al., 2014; Leszczynski, 2019; Wilmott, 2020).

2.4. Digital platforms and algorithmic constraints

2.4.1. Linguistic representation in databases

Digital mapping systems represent toponyms as discrete entries within structured databases, where each spatial entity is linked to a defined set of attributes, including a preferred linguistic form. This mode of representation requires a high degree of consistency and formal standardization, as database architecture relies on stable identifiers and unambiguous labels to ensure accurate retrieval and interoperability across systems. In practice, this means that each referent is typically associated with a single canonical form,

while alternative variants—whether orthographic, lexical, or context-dependent—are either subordinated or excluded (Graham et al., 2014; Kitchin et al., 2017).

From a linguistic perspective, this format imposes clear constraints on the representation of variation. The multiplicity of forms that characterizes toponymic usage in everyday contexts does not map easily onto database structures that require fixed entries and hierarchical organization. As a result, variation is often simplified, with one form designated as primary and others, if included at all, treated as secondary or searchable alternatives rather than equally valid options. This process transforms a variable linguistic system into a more rigid classificatory framework, in which diversity is reduced in favour of coherence and efficiency (Dodge & Kitchin, 2007; Graham et al., 2014).

The implications of this structuring extend beyond storage and retrieval. By determining which toponyms are displayed as default and which remain hidden or marginal, database design indirectly shapes patterns of exposure and use. Users are more likely to encounter and subsequently adopt the forms that are foregrounded within the system, while less visible variants become progressively less accessible. In this way, the technical requirements of digital representation contribute to the reconfiguration of toponymic practice, influencing not only how forms are organised, but also how they circulate in everyday language (Leszczynski, 2019; Shelton et al., 2015).

2.4.2. Logics of unification and disambiguation

Digital mapping systems represent toponyms as discrete entries within structured databases, where each spatial entity is linked to a defined set of attributes, including a preferred linguistic form. This mode of representation requires a high degree of consistency and formal standardization, as database architecture relies on stable identifiers and unambiguous labels to ensure accurate retrieval and interoperability across systems. In practice, this means that each referent is typically associated with a single canonical form, while alternative variants—whether orthographic, lexical, or context-dependent—are either subordinated or excluded (Graham et al., 2014; Kitchin et al., 2017).

From a linguistic perspective, this format imposes clear constraints on the representation of variation. The multiplicity of forms that characterizes toponymic usage in everyday contexts does not map easily onto database structures that require fixed entries and hierarchical organization. As a result, variation is often simplified, with one form designated as primary and others, if included at all, treated as secondary or searchable alternatives rather than equally valid options. This process transforms a variable linguistic system into a more rigid classificatory framework, in which diversity is reduced in favour of coherence and efficiency (Dodge & Kitchin, 2007; Graham et al., 2014).

The implications of this structuring extend beyond storage and retrieval. By determining which toponyms are displayed as default and which remain hidden or marginal, database design indirectly shapes patterns of exposure and use. Users are more likely to encounter and subsequently adopt the forms that are foregrounded within the system, while less visible variants become progressively less accessible. In this way, the technical requirements of digital representation contribute to the reconfiguration of toponymic practice, influencing not only how forms are organised, but also how they circulate in everyday language (Leszczynski, 2019; Shelton et al., 2015).

2.4.3. Visibility and lexical hierarchization

Not all toponyms are equally visible within digital systems, as their display is structured by ranking mechanisms, patterns of user activity, and commercial considerations. The ordering of results, the prominence of particular entries, and their recurrence across different interface elements are shaped by algorithmic criteria that privilege certain forms over others. Factors such as frequency of searches, volume of user contributions, and links to commercial entities all contribute to determining which toponyms are most readily accessible within the system (Shelton et al., 2015; Zook & Graham, 2007; Graham et al., 2014).

This uneven distribution of visibility gives rise to a hierarchical organization of toponymic forms. Some variants become dominant through repeated display and ease of retrieval, while others remain marginal, appearing only under specific conditions or not at all. The hierarchy is not necessarily aligned with local usage or linguistic relevance, but reflects the operational priorities of the platform, including efficiency, standardization, and economic value. As a result, forms that are widely used, institutionally recognized, or commercially salient are more likely to occupy central positions, whereas locally specific or less standardized variants are more likely to be peripheral (Graham & Zook, 2013; Light & Young, 2015).

The effects of this hierarchization extend to linguistic practice. Since users tend to rely on the most visible and accessible forms, repeated exposure reinforces the use of dominant toponyms and contributes to their stabilization in everyday discourse. At the same time, reduced visibility limits the circulation of alternative variants, which may gradually fall out of use. In this way, the distribution of toponyms within digital systems does not merely reflect existing patterns of usage but actively reshapes them by structuring the conditions under which different forms are encountered and adopted (Leszczynski, 2019; Wilmott, 2020).

3. RESEARCH METHODOLOGY

3.1. Research design

The study adopts a mixed-methods design that combines qualitative and quantitative approaches to account for both the internal structure of toponymic variation and the processes through which this variation is reshaped in digital environments. This methodological choice reflects the need to address toponyms simultaneously as linguistic forms, as elements of social practice, and as objects embedded in technological systems. A single-method approach would capture only partial aspects of this phenomenon: qualitative analysis is necessary to reconstruct meaning, usage, and indexical value, while quantitative modelling is required to identify patterns, measure distributions, and test relationships between variables across a larger population.

The research is structured as a sequential explanatory design, in which the qualitative component precedes and informs the quantitative stage. In the first phase, qualitative analysis is used to identify patterns of usage, distinctions between types of toponyms, and the meanings speakers attribute to different forms. This stage allows for the identification of relevant categories of variation—such as the contrast between standardized and vernacular forms, or differences in context-dependent usage—which are subsequently operationalized as variables. In the second phase, quantitative modelling is employed to test the distribution of these patterns across users and contexts, and to assess

the extent to which they are associated with factors such as frequency of platform use, degree of spatial familiarity, or type of locality. This sequential structure ensures that the statistical analysis is grounded in empirically observed linguistic behaviour rather than predefined assumptions (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 2010).

The design further allows for the integration of three complementary levels of analysis, each addressing a distinct dimension of the phenomenon under investigation. The first level is the discursive level, which focuses on how toponyms are used, interpreted, and negotiated in interaction. At this level, attention is given to patterns of variation in actual usage, the contextual conditions that influence the selection of particular forms, and the indexical meanings associated with different variants. This dimension is essential for understanding toponyms as elements of linguistic practice rather than as abstract labels.

The second level is the representational level, which concerns the way in which toponyms are encoded, organized, and displayed within digital platforms. Here, the analysis focuses on database structures, preferred forms, and the mechanisms through which certain toponyms are made more visible than others. This level makes it possible to examine how technical and algorithmic constraints shape the linguistic material available to users, and how these constraints translate into patterns of selection and standardization (Kitchin et al., 2017; Graham et al., 2014).

The third level is the behavioural level, which addresses the relationship between exposure to platform-displayed toponyms and their use in everyday language. This dimension connects representation and practice by examining whether and how the forms encountered in digital environments are taken up by users. It allows for the analysis of processes such as adoption, reinforcement, and potential displacement of alternative variants, and provides a basis for assessing the broader impact of digital mediation on linguistic behaviour (Leszczynski, 2019; Wilmott, 2020).

Taken together, this multi-level design makes it possible to move beyond a purely descriptive account of toponymic variation and to analyse the mechanisms through which such variation is structured, reduced, or transformed into digitally mediated contexts. By combining qualitative insight with quantitative testing and by linking discursive, representational, and behavioural dimensions, the study provides a comprehensive framework for examining the interaction between language, space, and platform-based systems.

3.2. Data and corpus construction

The empirical basis of the study consists of three interrelated datasets. The first dataset comprises a corpus of toponyms extracted from digital mapping platforms. The selection targets a defined set of urban locations, ensuring variation in terms of spatial scale, functional type (e.g. streets, landmarks, commercial entities), and degree of institutional codification. For each location, the platform-displayed toponym is recorded alongside any identifiable alternative forms, including official denominations and vernacular variants documented through external sources or field input.

The second dataset consists of semi-structured interviews with users. The sample is stratified to capture variation across patterns of platform use and spatial familiarity. Participants are grouped into categories such as regular users, high-frequency users (e.g. delivery workers, drivers), and users with strong local knowledge. The interviews focus on naming practices, recognition and recall of toponyms, perceived legitimacy of different forms, and reported usage in everyday interaction.

The third dataset includes quantitative survey data collected through a structured questionnaire. This component operationalizes key variables related to exposure, usage, and preference, allowing for statistical modelling of the relationship between platform interaction and toponymic behaviour.

3.3. Qualitative analysis: discursive and interpretative component

The interview data are analysed through thematic coding designed to capture both recurrent patterns and context-sensitive uses of toponyms. The coding scheme is developed through a combination of deductive and inductive procedures. On the one hand, an initial set of categories is derived from the theoretical framework, including distinctions such as standardised versus vernacular toponyms, indexical versus referential functions, and forms associated with institutional versus everyday usage. On the other hand, additional categories emerge from the data through iterative reading, allowing the analysis to remain sensitive to unanticipated patterns and to the specificity of speakers' accounts (Braun & Clarke, 2006; Guest et al., 2012).

The analytical process follows several stages. First, the interview transcripts are systematically reviewed in order to identify segments related to naming practices, recognition, and usage. These segments are then coded using an initial set of descriptive labels, which are subsequently refined into broader thematic categories. Coding is carried out iteratively, with constant comparison across cases to ensure consistency and to capture variation between speakers. Attention is given to instances in which speakers contrast different toponymic forms, justify their choices, or explicitly reflect on the perceived correctness or appropriateness of specific variants.

The analysis focuses on how speakers conceptualize and use toponyms within different communicative contexts. Central to this is the distinction between standardised forms—typically associated with official usage or digital platforms—and vernacular forms, which are embedded in local practices and informal interaction. Rather than treating this distinction as fixed, the analysis examines how speakers navigate between these categories, how they evaluate them, and how they deploy them in relation to audience and context.

Several analytical dimensions guide the interpretation of the data. The first dimension concerns the perceived authority of platform-displayed toponyms, including the extent to which speakers treat these forms as correct, legitimate, or preferable. The second dimension addresses patterns of variation in reported usage, focusing on the coexistence of multiple forms and the conditions under which one variant is selected over another. The third dimension examines the role of context, including factors such as interlocutor, communicative purpose, and degree of formality, in shaping toponym choice.

Also, the analysis explores how speakers position themselves in relation to different toponymic variants. This includes examining whether the use of certain forms indexes familiarity with a place, claims of local knowledge, or alignment with particular social groups. Attention is also given to cases in which speakers explicitly mark distance from certain forms, for example by attributing them to outsiders or to institutional usage. In this way, toponymic choice is treated as a resource for social positioning as well as spatial reference (Ainiala et al., 2012; Silverstein, 2003).

The qualitative component thus provides a detailed account of the meanings, functions, and distribution of toponymic forms in actual use. It allows for the identification of key variables—such as exposure to platform-based forms, degree of

reliance on digital tools, and patterns of alternation between variants—which are subsequently operationalized in the quantitative stage. By grounding the statistical analysis in empirically observed linguistic practices, this approach ensures that the modelling of toponymic usage reflects the complexity of the phenomenon rather than reducing it to predefined categories.

3.4. Quantitative analysis: operationalisation and measurement

The quantitative component is designed to capture, in a systematic and measurable way, both the degree of exposure to standardised toponyms within digital environments and the extent to which such exposure translates into their adoption in linguistic practice. The aim is not only to describe patterns of usage, but to model the relationship between platform-mediated input and actual toponymic behaviour across users and contexts.

The dependent variable is defined as the degree of alignment with platform-preferred toponyms. This construct is operationalized through a combination of self-reported usage measures and recognition-based tasks. In the usage component, respondents are asked to indicate which forms they typically use for a set of selected locations, allowing for the identification of preferences between standardised and vernacular variants. In parallel, recognition tasks assess familiarity with different forms by asking respondents to identify, match, or evaluate alternative toponyms. The combination of these measures makes it possible to distinguish between passive recognition and active usage, and to assess whether exposure leads to actual incorporation of standardised forms into speakers' productive repertoire.

Independent variables are selected in order to account for both behavioural and contextual factors that may influence toponymic usage. These include frequency of platform use (measured through reported daily or weekly interaction with digital mapping tools), degree of reliance on digital navigation (as opposed to memory or locally acquired knowledge), and level of familiarity with the spatial environment under consideration. Socio-demographic variables such as age, level of education, and, where relevant, occupational profile are also included, given their potential influence on patterns of technology use and linguistic behaviour. Together, these variables allow for the modelling of differences between user profiles and for the identification of factors associated with higher levels of alignment with standardised forms.

In addition to behavioural indicators, the analysis incorporates attitudinal variables designed to capture speakers' evaluations of different types of toponyms. These include perceived correctness, clarity, ease of use, and appropriateness in different contexts. Respondents are asked to rate alternative forms along these dimensions, providing insight into the normative frameworks that may support or resist the adoption of platform-preferred variants. This component is particularly relevant for understanding the extent to which standardisation is internalized as a linguistic norm rather than simply followed as a practical convenience.

To facilitate statistical analysis, composite indices are constructed for key dimensions of interest. An exposure index is developed to quantify the degree to which respondents encounter platform-displayed toponyms, combining measures of frequency of use, reliance on digital tools, and reported attention to map-based information. A usage index is constructed to capture the relative balance between standardised and vernacular forms in reported linguistic practice, based on responses across multiple items. These

indices are scaled and tested for internal consistency, ensuring that they provide reliable measures of the underlying constructs.

The resulting variables and indices are used as inputs in the statistical modelling stage, including multi-level analysis where appropriate. This allows for the examination of both direct effects (e.g. the impact of exposure on usage) and interaction effects (e.g. how this relationship varies depending on local familiarity or user profile). Through this approach, the quantitative component provides a structured and empirically grounded account of how standardised toponyms circulate and are adopted within digitally mediated environments.

3.5. Multi-level modelling (HLM)

To account for the hierarchical structure of the data, the study employs multi-level modelling. This approach allows for the simultaneous analysis of individual-level and context-level effects and is particularly suited to situations in which observations are nested within higher-level units.

At Level 1 (individual level), the model includes variables related to user behaviour and linguistic practice, such as frequency of platform use, exposure to standardised forms, and reported usage patterns.

At Level 2 (contextual level), variables capture characteristics of the spatial environment, including type of locality (e.g. large urban centre vs. smaller town), density of commercial entities, and degree of toponymic variation.

The model estimates both fixed effects (overall relationships between variables) and random effects (variation across contexts), allowing for the identification of cross-level interactions. This makes it possible to assess whether the influence of platform exposure on toponym usage varies depending on contextual factors.

3.6. Analysis of platform data

The dataset derived from digital platforms is analysed to identify patterns of standardisation and visibility. For each location, the platform-preferred toponym is compared with alternative forms, and discrepancies are recorded. Measures are constructed to capture the degree of alignment between platform representation and locally attested usage.

Additional indicators include the presence of commercial naming, frequency of updates (where accessible), and the extent to which multiple variants are represented or suppressed. This analysis provides an empirical basis for evaluating how platform structures affect the representation of toponyms.

3.7. Validity, reliability, and limitations

The study addresses validity through methodological triangulation, combining three distinct sources of evidence: qualitative interview data, quantitative survey data, and platform-derived toponymic data. Each source captures a different dimension of the phenomenon—discursive practices, reported behaviour, and digital representation—allowing for systematic cross-validation. Convergence between these sources is used to strengthen the interpretation of results, while divergences are treated analytically, as potential indicators of differences between reported usage, actual practice, and platform-mediated exposure (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 2010).

In addition, construct validity is supported through the explicit operationalization of key variables, such as exposure to platform-displayed toponyms and degree of

alignment with standardised forms, ensuring a clear correspondence between theoretical concepts and empirical measures. Reliability is addressed at multiple stages of the research process. In the qualitative component, coding is conducted using a structured coding scheme, with categories defined in advance and refined iteratively; where applicable, intercoder agreement procedures are used to ensure consistency. In the quantitative component, data collection follows standardised protocols, including uniform questionnaire design and controlled administration conditions. Composite indices (e.g. exposure and usage indices) are tested for internal consistency using standard reliability measures (e.g. Cronbach's alpha), and statistical analyses are conducted with attention to model specification, multicollinearity, and robustness checks. In the case of multi-level modelling, model fit and variance components are systematically evaluated to ensure the stability of estimates (Snijders & Bosker, 2012).

At the same time, several limitations must be acknowledged. First, access to platform algorithms and internal ranking mechanisms is restricted, which prevents direct observation of the criteria underlying the selection and prioritization of toponyms. As a result, inferences regarding platform effects are based on observable outputs rather than on the underlying computational processes. Second, the reliance on self-reported data introduces potential biases, including recall effects and discrepancies between reported and actual usage. Although the inclusion of recognition tasks partially mitigates this limitation, it cannot be fully eliminated. Third, the empirical focus on a defined set of locations and user groups limits the scope of generalization. While the selected cases are intended to capture variation across contexts, the findings should be interpreted as context-sensitive rather than universally applicable.

REFERENCES

AINIALA, Terhi, SAARELMA, Minna and SJÖBLOM, Paula, (2012), *Names in Focus: An Introduction to Finnish Onomastics*, Helsinki, Finnish Literature Society.

ALDERMAN, Derek, (2008), "Place, naming, and the interpretation of cultural landscapes", in Brian GRAHAM & Peter HOWARD (eds.) *The Ashgate Research Companion to Heritage and Identity*, Aldershot, Ashgate, pp. 195-213.

AZARYAHU, Maoz, (2011), "The critical turn and beyond: The case of commemorative street naming", in *ACME: An International Journal for Critical Geographies*, 10(1), pp. 28-33.

BERG, Lawrence D., & VUOLTEENAHO, Jani, (2009), *Critical Toponymies: The Contested Politics of Place Naming*, Farnham, Ashgate.

BRAUN, Virginia, & CLARKE, Victoria, (2006), "Using thematic analysis in psychology", in *Qualitative Research in Psychology*, 3(2), pp. 77-101.

CRESWELL, John W., & PLANO CLARK, Vicki L., (2018), *Designing and Conducting Mixed Methods Research*, 3rd edn, Thousand Oaks, SAGE.

DODGE, Martin, & KITCHIN, Rob, (2007), "Rethinking maps", in *Progress in Human Geography*, 31(3), pp. 331-344.

FARMAN, Jason, (2012), *Mobile Interface Theory: Embodied Space and Locative Media*, New York, Routledge.

GRAHAM, Mark, & ZOOK, Matthew, (2013), "Augmented realities and uneven geographies: Exploring the geolinguistic contours of the web", in *Environment and Planning A*, 45(1), pp. 77-99.

GRAHAM, Mark, ZOOK, Matthew & BOULTON, Andrew, (2014), "Augmented reality in urban places: Contested content and the duplicity of code", in *Transactions of the Institute of British Geographers*, 38(3), pp. 464-479.

GUEST, Greg, MACQUEEN, Kathleen, & NAMEY, Emily E. (2012), *Applied Thematic Analysis*, Thousand Oaks, SAGE.

HAUGEN, Einar, (1966), "Dialect, language, nation", in *American Anthropologist*, 68(4), pp. 922-935.

HOUGH, Carole, (2016), *The Oxford Handbook of Names and Naming*, Oxford, Oxford University Press.

JORDAN, Peter, (2012), 'Place names as ingredients of space-related identity', Names, 60(1), pp. 1-9.

JOSEPH, John E., (1987), *Eloquence and Power: The Rise of Language Standards and Standard Languages*, London, Frances Pinter.

KADMON, Naftali, (2000), *Toponymy: The Lore, Laws and Language of Geographical Names*, New York, Vantage Press.

KITCHIN, Rob, LAURIAULT, Tracey P., & WILSON, Matthew W., (2017), *Understanding Spatial Media*, London, SAGE.

LESZCZYNSKI, Agnieszka, (2019), "Digital methods III: The digital mundane", in *Progress in Human Geography*, 43(3), pp. 544-553.

LIGHT, Duncan, & Young, Craig (2015), "Toponymy as commodity: Exploring the economic dimensions of urban place names", in *Environment and Planning A*, 47(2), pp. 435-450.

MILROY, James, & MILROY, Lesley, (2012,) *Authority in Language: Investigating Standard English*, 4th edn., London, Routledge.

MILROY, James, (2001), "Language ideologies and the consequences of standardization", in *Journal of Sociolinguistics*, 5(4), pp. 530-555.

OCHS, Elinor, (1992), "Indexing gender", in Alessandro DURANTI & Charles GOODWIN (eds.) *Rethinking Context: Language as an Interactive Phenomenon*, Cambridge, Cambridge University Press, pp. 335-358.

ROSE-REDWOOD, Reuben, ALDERMAN, Derek H., & AZARYAHU, Maoz, (2010), "Geographies of toponymic inscription: New directions in critical place-name studies", in *Progress in Human Geography*, 34(4), pp. 453-470.

SHELTON, Taylor, ZOOK, Matthew, & WIIG, Alan, (2015), "The "actually existing smart city"", in *Cambridge Journal of Regions, Economy and Society*, 8(1), pp. 13-25.

SILVERSTEIN, Michael, (2003), "Indexical order and the dialectics of sociolinguistic life", in *Language & Communication*, 23(3-4), pp. 193-229.

SNIJDERS, Tom A. B., & BOSKER, Roel J., (2012), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edn., London, SAGE.

TASHAKKORI, Abbas, & TEDDLIE, Charles, (2010), *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, 2nd edn., Thousand Oaks, SAGE.

TENT, Jan, & BLAIR, David, (2011), *Motivations for Naming: The Development of a Toponymic Typology for Australian Placenames*, Adelaide, ANPS.

WILMOTT, Clancy, (2020), "Small moments in spatial big data: Calculability, authority and interoperability in everyday mobile mapping", in *Big Data & Society*, 7(2).

ZOOK, Matthew, & GRAHAM, Mark (2007) "Mapping DigiPlace: Geocoded internet data and the representation of place", in *Environment and Planning B*, 34(3), pp. 466-482.