Names | A Journal of Onomastics

Mapping Place Names

Søren Wichmann
Kiel University, GERMANY

Lennart Chevallier
Kiel University, GERMANY

ans-names.pitt.edu

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

Articles in this journal are licensed under a Creative Commons Attribution 4.0 International License.

This journal is published by Pitt Open Library Publishing.

Abstract

This paper demonstrates how to leverage the GeoNames data for seeking patterns in toponymic data using the software package 'toponym', which we wrote for the R computational environment. After discussing a distinction between particularistic and pattern-seeking approaches to toponymics, we go on to characterize the data of GeoNames, which are particularly appropriate for the latter type of approach. Then, we present two cases studies. The first case study is on Xincan place names in Guatemala, and the second is on Slavic place names in eastern Germany. These explorations support our hypothesis that toponymics may benefit from new computational tools.

Keywords: toponym, hydronym, R, software, GeoNames, Xincan, Slavic

General Introduction

Place names (toponyms) are perceived as potentially imbued with meaning (Radding & Western 2010; Algeo & Algeo 2000), yet we also accept the apparently arbitrary relationship between a designation and its referent as we encounter it. Thus, newcomers in a region often adopt existing place names, even if they are opaque to them, and toponyms can become important sources of information that "permit historical inferences about languages and the people who spoke them" (Campbell 2013, 436).

For instance, the US states of Mississippi (Baca 2007) and Alabama (Read 1994) are teeming with Choctaw names, bearing witness to parts of the extension of this ethnic group before most Choctaws underwent forced relocation to Oklahoma in 1831–33. As the editor of the Vicksburg Daily Sentinel, eyewitness to the removal, wrote, "They leave names to many of our rivers, towns, and counties, and so long as our State remains the Choctaws who once owned most of her soil will be remembered" (DeRosier 1970, 4). Today only native speakers and linguistic specialists will be able to interpret the meaning of old Choctaw place names, for instance (in Mississippi, following Baca 2007) *Tchoutacabouffa* 'Broken Pottery' *River*, *Itta Bena* 'Forest Camp', *Tallahoma* 'Red Rock' *Creek*, and (in Alabama, following Read 1994) *Buttahatchee* 'Sumac River', *Coatopa* 'Wounded Panther Creek', *Nanafalia* 'Long Hill', and so on. Even for specialists, the majority of names of likely Choctaw origin remain more or less obscure.

The opposite case, where new names are given to locations already carrying names in a language other than the one of the newcomers, is of course also common. Sometimes this is done systematically for political motives. For instance, in 757 CE King Kyeongdeok of Silla changed toponyms in the realm of his kingdom (within modern-day North and South Korea) into standardized names written with two Chinese characters (Endo 2021, 35). Closer to the present, in the years after 1945, hundreds of place names in formerly German-occupied areas of Czechoslovakia were renamed from German into Czech (Lehmann 1999). Finally, during 1949–1960, the vast majority of Arabic place names were replaced with Hebrew ones in the new state of Israel (Azaryahu & Golan 2001), a practice which continued as new territories were included after the Six Days War of 1967 (Cohen & Kliot 1992). Interestingly, in all three cases translating the place names was a common, consciously applied strategy. This testifies to the conservative power inherent in toponyms, even in the face of radical reformative stances.

Detailed philological work on place names commonly involves seeking out documentation for the earliest attestations (cf. already Petersen 1833) and carrying out lexical, morphological, and phonological analysis (e.g., Bright 1998). In addition to this line of work, onomastic research may profit from the inspection of geographical distributions of crude phoneme/letter sequences. For instance, linguistic work tells us that the suffix *-ham* contained in many British place names is West Germanic and means something like 'dwelling, homestead, hamlet, for example.', and the study of the documentary record indicates that it begins to appear in the 5th century (Copley 1988, 31). Even without linguistic analysis and archival work, however, we can hypothesize just from the distribution of this string that *-ham* is a suffix with a relatively generic meaning belonging to a stratum of English place names which is neither of Celtic nor of Scandinavian origin. Figure 1 shows 817 contemporary occurrences of this string in Great Britain and Ireland. It is infrequent in Scotland, Wales, and Ireland, and not in any way confined to the Danelaw area. Clearly the gazetteer of Copley (1988, 31–35), which discusses nine early place names in *-ham*, is well complemented by a raw distributional map of place names ending in the letters *-ham*, even if the latter does not directly tell us anything about the meaning of the apparent suffix and contains no historical stratification.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

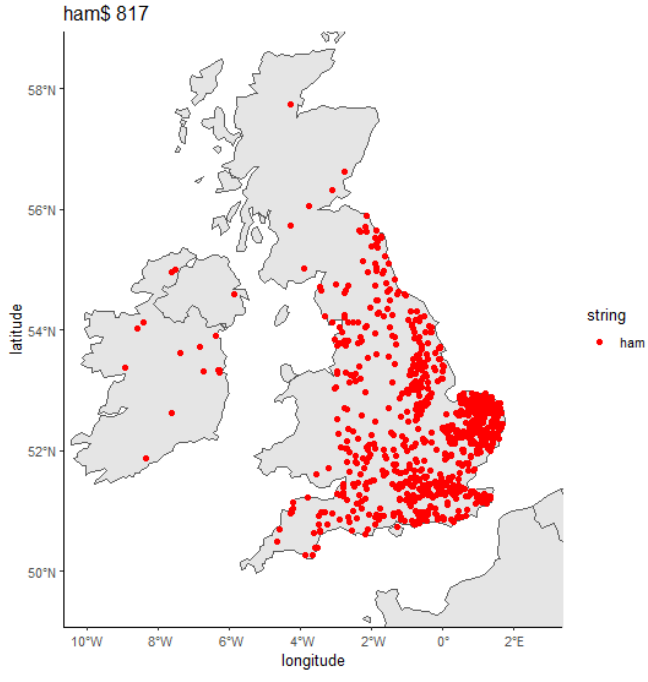

Figure 1: The Distribution of *-ham* in the United Kingdom and Ireland as Displayed Using the Toponym Package

With the advent of the personal computer it has become easier to generate distributional maps of place names containing particular strings. An early example is the overview of common terms for "stream" (*creek*, *branch*, *run*, *brook*, and so on) in Campbell (1991), which draws upon the Geographic Names Information System (GNIS) of the United States.[1] The author acknowledges technical assistance for producing the maps. The introduction of geographic information systems (GIS) has made it easier to produce custom digital maps, and publications in toponymy have profited from these technological advances. Chen et al. (2014, 6,315–6,316) briefly reviews many early GIS-based toponymic studies. Among later studies we can mention Fuchs (2015), which employs ArcGIS in the plotting and distributional analysis of German place names in the US Midwest and Endo (2021) which, using ArcGIS online, plots Korean place names from the first millennium CE with different linguistic backgrounds (Japonic, Koreanic, Sinitic, Tungusic, Mongolic).

In the preceding we have exemplified two general approaches to toponymy—a mainly philological approach, focusing on individual names, including their etymologies, and a pattern-seeking approach, simultaneously considering multiple names. Obviously the two types of approach are not sharply distinct, nor do they reflect some kind of opposition—instead, they are better characterized as complementary. Tent (2015) suggests distinguishing them terminologically. This author discusses contrastive terms that emphasize different facets of the two different approaches, including "micro" vs. "macro", "case" vs. "pattern" analysis, and "qualitative" vs. "quantitative" research, but his preferred terms are "intensive" vs. "extensive" toponymic research. We would concur with this terminology, although other candidates might be "particularistic" (a term also used by Hunn 1996, 4) vs. "pattern-seeking". As for the terms 'qualitative' vs. 'quantitative' for the two approaches, also employed by Tent (2015), we would argue that a quantitative approach is one in which numbers and statistics matter, not simply one that draws upon many data-points. Thus, when Tent (2015, 67) presents a table according to which the quantitative approach is used in 15.6% of papers in the journal *Names* in 1952–2014, with the rest representing qualitative, mixed or other approaches, we suspect that the percentage of papers that are truly quantitative in nature is actually smaller. It is in any case clear that there is plenty of room for applications and developments of the pattern-seeking approach to toponomy, and even more room for the introduction of statistical methods. We see our own work as a contribution to the pattern-seeking approach to toponymy, and our hypothesis in this paper is that new tools aiding research using this type of approach have a great potential.

The present paper illustrates the use of a package written in the R computational environment (R Core Team 2022) which is designed to filter and display any subset of millions of place names from across the

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

world, producing maps like the one in figure 1 and allowing for outputting tables that include relevant names and corresponding geographical coordinates. We offer this tool for the science of onomastics in particular and also for any other purpose for which it might prove useful. Onomastics, as is well known, is not restricted to place names, but its branch toponymy is. Subbranches hereof include hydronomy (the study of names of rivers, lakes or other bodies of water), oronymy (the study of the names of hills, mountains or mountain ranges), speleonymy (the names of caves, chasms, grottoes, mines, and so on.), and hodonymy (the names of streets, avenues, lanes, and so on.) (Perono Cacciaforo & Cavallaro 2023, 3; Coates 2013). All these subbranches of toponymy are served at least to some extent by our package. In the following sections, we first introduce the data that the package draws upon and finally we provide two case studies illustrating the utility of the tool. The nuts and bolts of the package itself are presented in a separate Note (Chevallier & Wichmann 2024). An online text file[2] that follows the outline of the paper repeats all lines of code in the text and also contains additional code allowing for figures to be recreated and results to be fully replicated.

Methodology

The tool used to carry out toponymic research in this paper is the R package 'toponym' (see Chevallier & Wichmann 2024 for a practical introduction). The package provides an interface to the GeoNames data. Here we provide an overview of the contents of this database, which is hosted at https://www.geonames.org/. GeoNames (GeoNames 2023) is an evolving database of public domain geographical data. The version of the database downloaded for this paper, as embodied in the allCountries.txt file (last modified at 03:52, November 17, 2023), contains 12,492,864 entries pertaining to 253 countries and 9 feature classes. The information is aggregated from more than 400 sources[3] and is entirely free for download. Figure 2 shows the distribution of entries over the different feature classes, which are defined in the caption to the figure. As is clear from figure 2, around two thirds of the database consist of names for the built or humanly defined environment (P, S, A, L, R; ~8.5 million entries), with features of the landscape (H, T, V, U; ~4.0 million entries) constituting the remaining third.

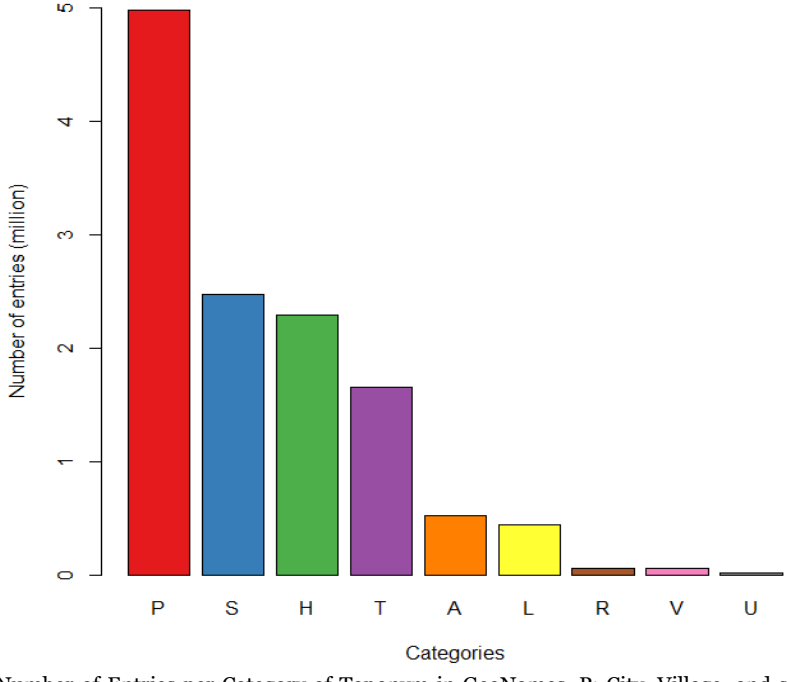

Figure 2: Number of Entries per Category of Toponym in GeoNames. P: City, Village, and so on; S: Spot, Building, Farm; H: Stream, Lake, and so on; T: Mountain, Hill, Rock, and so on; A: Country, State, Region, and so on; L: Parks, Area, and so on; R: Road, Railroad; V: Forest, Heath, and so on; U: Undersea

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

We now briefly describe the columns of the GeoNames database, referring to column names in square brackets. Of the entries, 100% contain an integer ID [geonameid], a toponym [name], geographical coordinates [latitude, longitude], and a modification date [modification date]; 95.33%–99.99% of the entries additionally contain an ascii version of the toponym [asciiname], a feature class [feature class] (see the caption to figure 2), a feature code [feature code] indicating a subclass of the feature class, a two-letter country code [country code], a code for highest-level administrative units (such as states in the US or Germany) [admin1 code], a population figure [population], a digital elevation model (in meters) [dem], and a time zone [timezone]; and 53.63%–67.20% of the entries contain alternate names [alternatenames] and codes for second-level administrative units [admin2 code]. Finally, 00.10%–20.86% of the entries contain third- and fourth-level [admin3 code, admin4 code], elevation (in meters) [elevation], and alternate country conde [cc2].

In a few cases the contents of a column is not self-explanatory, calling for a few comments. The [name] column contains toponyms in the standard spelling whenever the relevant writing system is Latin based. When the writing system is not Latin based, the column will feature a Latin-based transcription and the name rendered in the national orthography will instead normally be found in the [alternatenames] column—possibly alongside other transliterations and/or variants. Non-Latin orthographic renderings of toponyms, however, are provided less systematically than their counterpart Latin transcriptions. For instance, 11.1% of all entries for Russia have empty cells in the [alternatenames] column, which holds Cyrillic versions. Similarly 4.4% of entries for China have empty cells in [alternatenames]. Thus, more complete results are expected when accessing [name] than [alternatenames]. The relationship between the [dem] and the [elevation] columns also invites comments. The [dem] column contains srtm3 (Shuttle Radar Topography Mission 3) or gtopo30 (Global 30 Arc-Second Elevation) data, but the documentation does not specify the origin of the [elevation] column data. The values are not identical, but correlate as $r = .997$ by a Pearson correlation, which shows that they measure the same quantity but also that there is room left for large discrepancies in individual cases. More importantly, only 19.14% of the entries have filled cells for [elevation], whereas [dem] is near-complete. Hence, it would normally make little sense to use [elevation]. The GeoNames documentation[4] provides additional detail on the contents of the database. As far as we are aware, there is no other single database of place names that might compete with GeoNames as a source of toponymic research. Acheson et al. (2017) offer comparisons with the next largest global gazetteer, the Getty Thesaurus of Geographic Names (TGN 2023), noting that GeoNames (at their time of writing) was about ten times as large.

Results

Tracking the Former Extent of Xincan Languages

In a book on Maya culture, Thompson (1970, 98–99) suggested that place names in eastern Guatemala and western Honduras ending in *-agua* or *-ahua*, or their reduced forms *-gua* or *-hua*, may relate to a particular non-Maya ethnic group which got partly displaced by the Ch'orti' Mayas but had a stronger presence in eastern Guatemala before the latter arrived on the scene. Thompson was not able to identify the ethnic group more precisely. His presentation of the relevant toponymic data is worth quoting *in extenso* since it provides for an interesting comparison with the output of the 'toponym' package

> In eastern Guatemala and western Honduras, approximately the area of the Chorti at the time of the Conquest, are Toquegua Indians, Motagua and Managua rivers; and the sites of Tigua north of Camotan, Tigua east of Paraiso, Chanmagua or Chaumagua, Jagua, Pasasagua, Anchagua, and Cocuyagua.
>
> In central Honduras, approximately between 86° 50' and 88° 10', that is to say, east of Chorti territory, are the rivers Jacagua, Jalagua, Chilistagua, Comayagua, Sasagua, Chasnigua, and perhaps Ulua; and the sites of Silisgualagua, Manzaragua, Masahua, Mulacagua, Tircagua, Chumbagua, Chapulistagua, Xagua, Xelegua, Eraxagua, Moncagua, Teconalistagua, Laxigua, Talgua, Colomoncagua, Tiscagua, Apacilagua, Conchagua, as well as Masaguara, Sicaguara, Yaguacire, and Yamaguare.
>
> In El Salvador we find Comasahua, Atepammasagua, Quixnagua, Masahua, and Moncagua, as well as Aguasarca and Guahtajigua.
>
> One may also note other examples in territory even farther east and southeast, such as Nicaragua, Managua, and even Veragua. (Thompson 1970, 98–99)

It can be verified that the majority of the particular place names mentioned are actually in GeoNames by entering a search for the specific names. The result is displayed in figure 3. It shows 29 of the 49 place names.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

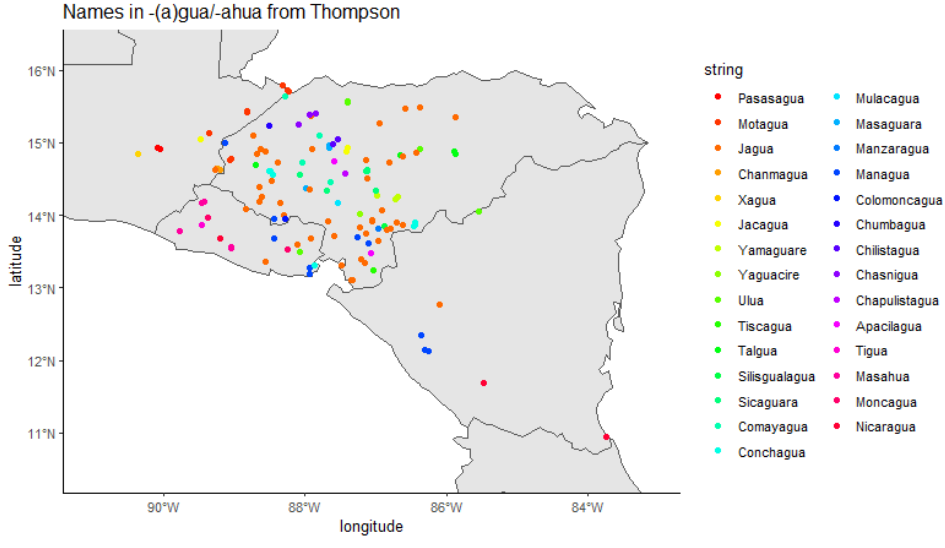

Figure 3: Attempt to Map All the Specific Place Names Hypothetically Associated with the "agua people" in Thompson (1970, 98–99)

To replicate Thompson's findings starting out from recurrent substrings rather than full names, the strings to search for are final *agua*, *gua*, *ua*, *ahua*, *guara*, *guacire*, *guare*. Searching for these among names for populated places and hydronyms in Guatemala, Honduras, El Salvador, and Nicaragua could, in principle, return all the names mentioned by Thompson except *Toquegua*, which is an ethnonym, and *Aguasarca*, where *agua* appears initially—which makes its inclusion somewhat dubious anyway. Additional names are expected to appear. The output data frame contains 477 rows, but in many of these the target string pertains to Spanish words: *antigua* 'ancient', *legua* 'league', *yegua* 'mare', *Padua* 'Padua', *jagua* 'species of tree (Genipa americana)' [the last was included in Thompson's list although the *jagua* is well-known name of a tree in several varieties of Latin American Spanish]. After filtering we are down to 181 rows. The coordinates of the revised data frame are plotted in figure 4.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

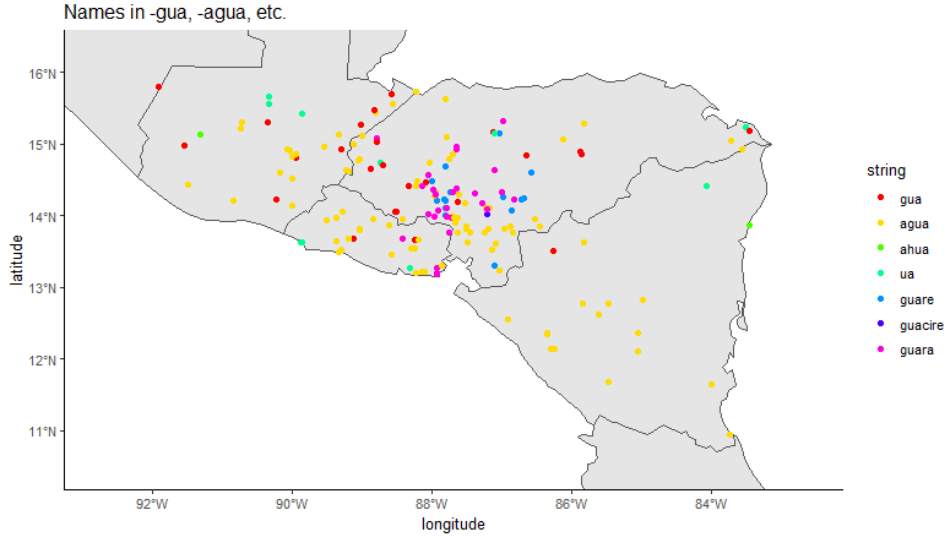

Figure 4: Trailing Strings *agua*, *gua*, *ua*, *ahua*, *guara*, *guacire*, *guare* in Toponyms in Guatemala, Honduras, El Salvador, and Nicaragua, with Spanish Forms Removed

The maps in figures 3 and 4 greatly help to evaluate Thompson's suggestions. Regarding the general distribution of *a(gua)* names, the main insights from the extended search (figure 4) are that more of southern Guatemala should be included than is apparent from Thompson's specific examples. We also more clearly see a radial distribution roughly centered on 14°N, 88°W, rather far from the area around the border between Guatemala and Honduras that was Thompson's concern.

Campbell (1978) cites the earlier suggestion of Thompson (1970, 98–99), but brings other relevant parts of toponyms to the table, and his focus is on a historically known ethnic entity, namely Xincan. Xincan is a small family of some four to five languages formerly spoken in southeastern Guatemala. Campbell's latest extensive discussion is in his 1998 publication, which is the study we will refer to here. The same ideas are consistently repeated in multiple publications (Campbell 2013, 1998, 1978) as well as in a recent talk (Campbell 2022), slides for which were generously shared with us by the author.

As in our handling of the data from Thompson (1970), we can first map specific names given by Campbell (1998, 188) and, after that, the distribution of some specific strings supposedly pertaining to Xincan toponyms. All the specific names mentioned by Campbell except four (*Sanguayaba*, *Cerro Sansuque*, *Sanyoyo*, and *Arloroma*[4]) are found in the database, and we can provide a plot that was never provided by Campbell but would represent his view (see figure 5). The map shows a compact area in southeastern Guatemala, where the border seen is the one shared with El Salvador. This would seem to be an adequate representation of Campbell's view of the historical extension of Xincan as revealed through toponyms. In the 1998 paper he describes the area as extending "from the Motagua Valley to the Pacific Coast and from near Guatemala City [e.g., San Pedro Ayampuc] to Honduras and El Salvador",[5] and in the 2022 talk there is an explicit statement that Xincan never extended into (what is currently) El Salvador.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

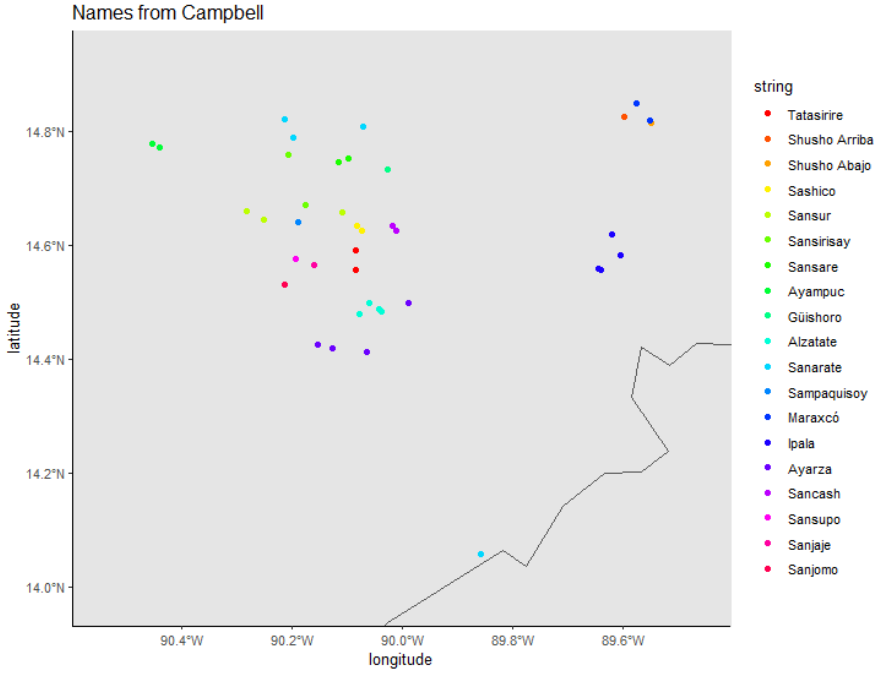

Figure 5: Specific Names Associated with Xincan According to Cambell (1998, 188)

Up to this point we have neglected some additional evidence that Campbell cites, namely some *agua*-type names that Thompson (1970) had also discussed. Following the excerpt given above Campbell writes:

> Thompson was on the right track; many of his examples turn out to be from Xinca *šawi̱ɬi* 'town' (ultimately derived from the verb root *šawi̱* 'to dwell'). [. . .] The following (many from Thompson's list of "*-agua*" towns) appear to incorporate this Xinca term for 'town': Pasasagua (compare Xinca *ṣan-paṣa?* 'Pasaco' [a town]), *Jagua*, *Sasagua* (Xinca *ṣa(n)-šawi̱* 'in-town'), Xagua, Eraxagua (*i̱ra* 'big', i.e. 'big town'), Conchagua, Comasahua, Manzaragua, Anshagua, Anshigua, Xororagua, etc. None of these has a Mayan or Nahuatl source, and the Xincan origin is clear for most. Campbell (1998, 188-189)

When we try to map these names the result is as in figure 6. As it turns out, there are multiple instances of *Conchagua* in Honduras and El Salvador, but apparently none in Guatemala; *Manzaragua* is only in Honduras; and *Jagua* is widespread, but this is a Spanish tree term as mentioned above, so we can simply ignore that. The results in figure 6 suggest that names in *-shigua*, *-shagua*, *-xagua,* and (sometimes) *-sagua*, but not those in *-chagua* and *-ragua* are good candidates for pertaining to Xincan, unless we want to accept an extension way beyond southeastern Guatemala for which evidence is otherwise lacking.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

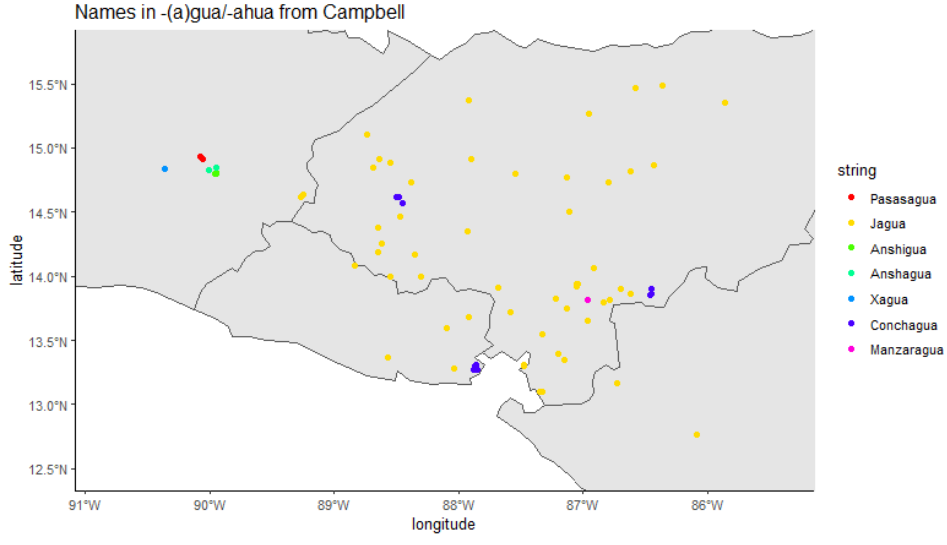

Figure 6: Specific Place Names Associated with the "agua people" in Thompson (1970, 98–99) and Taken by Campbell (1998) to be Xincan

While the proposal of Thompson was pattern-seeking and that of Campbell mainly particularistic we hope to have demonstrated that our software package was helpful in investigating both proposals through the maps produced. Although a few specific names should be removed from Campbell's list of Xincan toponyms his proposal could be supported. [6] As for Thompson's "*agua*" people it seems unlikely that the data supports a hypothesis linking a toponymic ending similar to *-agua* to a specific ethnic entity, although the idea merits further investigation. We now know more about the distribution of the toponyms containing the endings that Thompson discussed, which is an essential first step in the further investigation of this proposal.

Slavic Toponyms in Germany

In this subsection we demonstrate how our package can be helpful in providing distributional overviews of the most common Slavic toponymic elements in German place names. They pertain to West Slavic varieties spoken in the eastern parts of what is now Germany before Germanic populations expanded eastward into these regions causing assimilation of the Slavic populations and, to a large extent, loss of the Slavic languages there.

It is with some trepidation that we approach the topic of place names of the formerly Slavic parts of Germany, since this is one of the best—if not *the* best—researched geographical areas with respect to toponyms in the entire world. This strand of research is part of a long tradition of toponymics in Germany, a tradition which was already established by the middle of the 19th century (Förstemann 1863), and it flourished particularly within the German Democratic Republic from 1949 to 1990 (Walther 1995), also continuing vigorously after German reunification (Bily 2002; Eichler 1994-95). While it is challenging to confront such a well-researched topic, the former Slavic area of Germany is an attractive target for toponymics because here place names are the best and sometimes the only source of information about language distributions in past centuries (Eichler & Witkowski 1985b, 51). Slavic groups moved into the thinly populated area from the 6th century onwards (Bily 2003, 22; Plate 1998, 14–17). In the well-known classification of Slavic languages into East, West, and South Slavic, these groups belong to West Slavic, but in the historical sources they are referred to by a bewildering number of different names and are often referred to as different "tribes" in the literature (Hermann 1985). It appears that there is not a one-to-one match between ethnicity and language, so ethnic designations are of relatively limited use in a linguistic

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

classification in this case. Sometimes they are simply designations derived from the area dominated by a certain tribe (Eichler 1998, 276–277; Eichler & Witkowski 1985a; Hermann 1985, 9).

While we would like to develop a clear, data-driven, well supported, and tractable model for the erstwhile distribution of Slavic varieties in Germany, it turns out that a blunt pattern-seeking approach will not capture borders between Slavic varieties because differences are too subtle. Instead, some distributional differences in toponyms that are due to German dialect differences tend to emerge. We will, however, at least be able to delineate the extension of the former Slavic territory.

The suffix that is both most widely and most densely distributed is *–itz* (see figure 7). Some of the cases of the trailing string *-itz* belong to *-witz*, which is also Slavic. They originate in Slavic suffixes *-ici*/*-ovici* or *-ica*/*-ovica* (cf. Eichler 1976, 137). Not shown in figure 7, which is restricted to Germany, is the northernmost extension represented by a handful of placenames in *-itse* on the southern islands of Lolland and Falster in Denmark (Thorndal 1963). Outliers to the west mostly pertain to German names ending in a string *-itz* which is not identical to the Slavic suffix, such as *Titz* (North Rhine-Westphalia), *Merkenfritz* (Hesse) or *Olpenitz* (Schleswig-Holstein). But a closer look at each may reveal evidence for some isolated Slavic settlements outside of the general Slavic area.

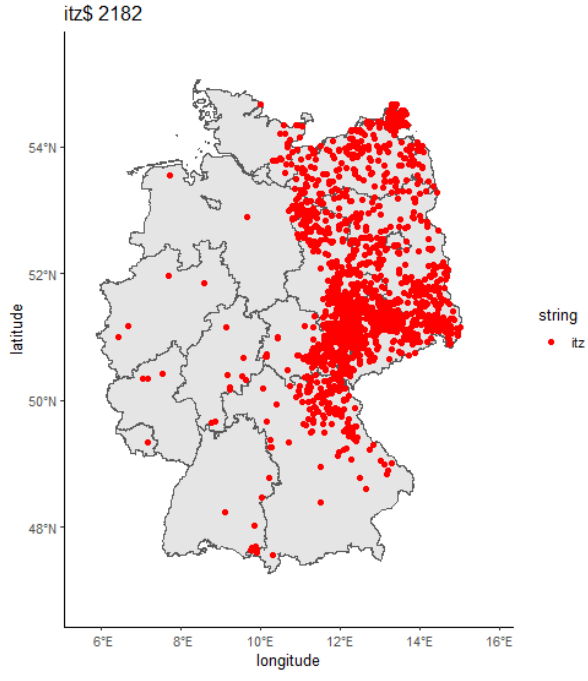

Figure 7: Distribution of Names of Populated Places in Germany Ending in *-itz*

Another typical Slavic suffix is *-ow*. Its distribution is mapped in figure 8. This originates in a Slavic suffix *-ov* or *-ava* and may derive place names from either common nouns (appellatives) or proper (personal) names. Typical shapes reflecting *-ov*/*-ava* in Low German are *-ow* and *-o*, while Middle and High German may show *-au*, *-a*, *-e* or zero, according to Eichler (1976, 139). Thus, the distribution is somewhat deceptive because the Slavic suffix is present throughout the "itz-area" in other guises even if the shape *-ow* does not continue further south into Sachsen and Bayern. It is an example of German dialectal variation which interferes with phonological shapes of the attested Slavic toponyms.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

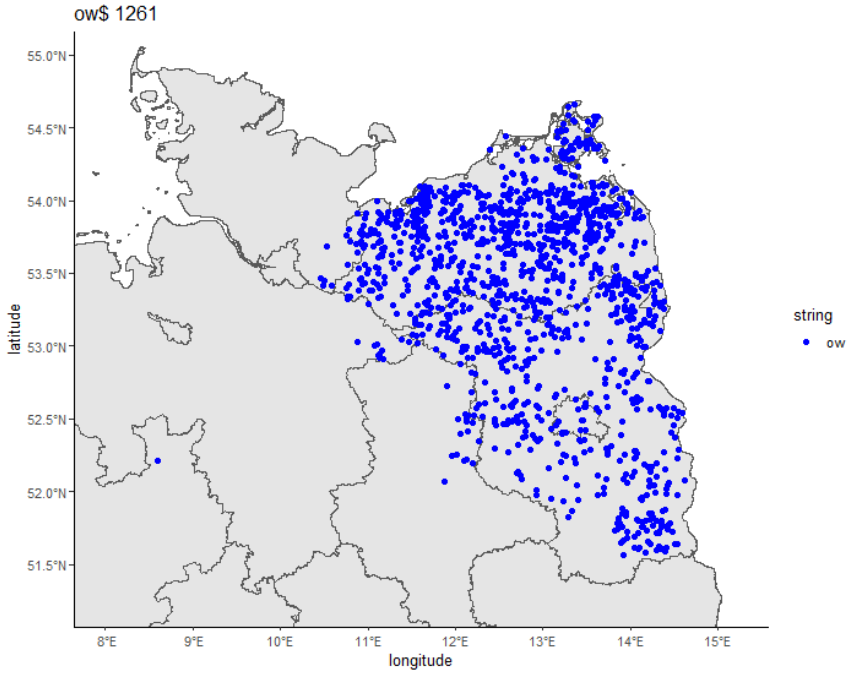

Figure 8: Distribution of Names of Populated Places in Germany Ending in *-ow*

Interference from German dialectal variation also affects the third of the most common Slavic suffixes, namely *-in*. This originates in the Slavic suffixes *-in*, *-ina*, *-yn, and –ɔn*, and is reflected as *-in* or *-en* in Low German and *-en*, *-n*, *-a*, *-e* in Middle and High German according to Eichler (1976, 13). In figure 9 we map the distribution of *-in*, with only a little filtering, removing cases where *-in* belongs to some obviously non-Slavic names. As in the case of *-ow*, the distribution of the original Slavic suffix is expected to cover the whole "itz-area", but we would have to analyze hundreds of toponyms individually to find the more extended distribution—maps of *-en*, *-n*, *-a*, or *-e* without further filtering produce dots all over Germany.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

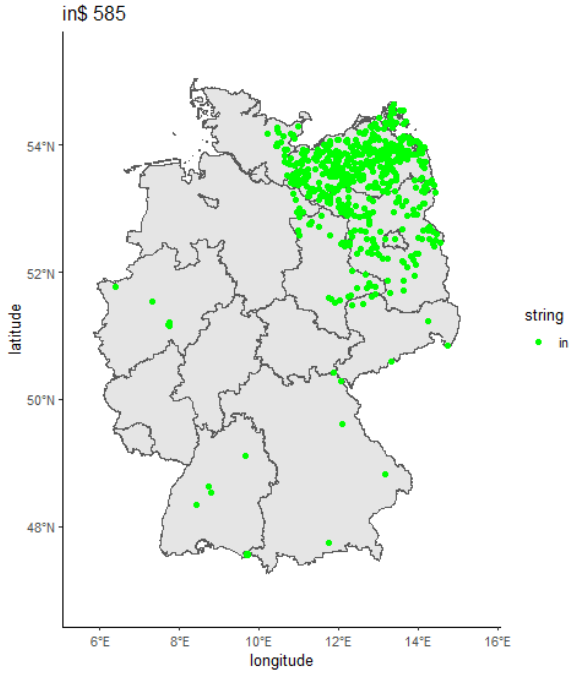

Figure 9: Filtered Distribution of Names of Populated Places in Germany Ending in *-in*

After the establishment of Slavic-German sister towns due to the eastward spread of German populations, modifiers such as *Groß-/Klein-*, *Alt-/Neu-*, *Nieder-/Ober-*, and *Wendisch-/Deutsch-* were applied to distinguish the two (Eichler and Walther 2001, Vol. 1, xxxiv). In the case of *Groß-/Klein-* it was the Slavic town that received the modifier *Klein-* (Bily 2015, 133; Gringmuth-Dallmer 1990). All pairs of modifiers except *Wendisch-* vs. *Deutsch-* are used in other contexts as well and will not help us to delineate the Slavic area. In figure 10 we map *Wendisch-* vs. *Deutsch-*. *Wendisch* is a generic term which roughly translates to "Slavic". Since we are only interested in the string *Deutsch-* when it is potentially paired up with *-Wendisch*, we do not include states where only *Deutsch* is found. And in order to better focus on the area of interest we exclude from the map an unexplained single occurrence of *Wendisch-* in Baden-Württemberg in southern Germany. Although thinly distributed, *Wendisch-* does roughly bring out the contours of the Slavic area. Additionally, the map shows that *Wendisch-* can occur both alone and paired with *Deutsch-* (one pair in Lower Saxony, one pair in Saxony-Anhalt, at least two pairs in Saxony).

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

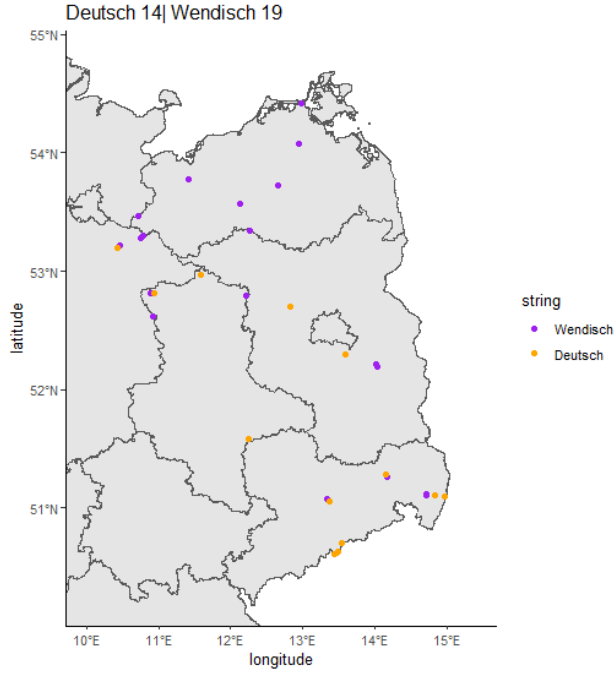

Figure 10: *Wendisch-* and *Deutsch-* in States Where *Wendisch-* Is Found (Minus Baden-Württemberg)

In figure 11 we combine the maps showing the three most widely occurring Slavic suffixes as they can readily be identified in German place names, as well as a map of the modifier *Wendisch* (now without *Deutsch*). The area in eastern Germany where the dots are densely distributed corresponds to the former Slavic territory, whereas the more scattered outliers to the west are either individual incursions or, probably more commonly, due to strings not pertaining to Slavic suffixes. There are, to our knowledge, not any Slavic suffix strings with frequencies anywhere near those of *-itz*, *-ow*, and *-in*. A few could be added which are also found throughout the former Slavic area, but they would be too thinly distributed to change the general picture. An example is *-ehna*, which is characteristic of some Slavic place names according to Eichler (1976, 133) and which occurs widely, albeit sparingly.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

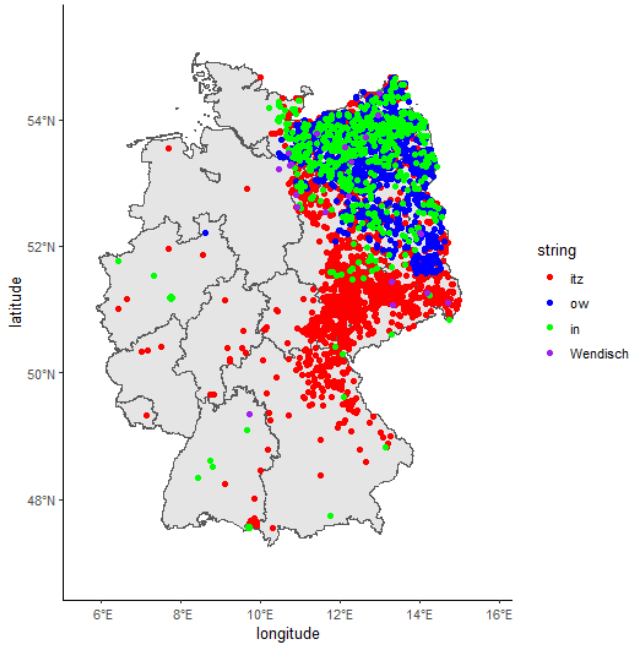

Figure 11: Distribution of *-in* (filtered), *-itz*, *-ow*, and *Wendisch-*

We tried to find isoglosses providing evidence for linguistic subdivisions within the Slavic area. There is a known major division between Polabian in the north and Sorbian in the south running about midway through the area. The border roughly coincides with the southernmost extension of the "in-area" in figure 9 (see the map in Eichler & Witkowski 1985b, 52; and cf. further Bathe et al. 1971; Fischer 1970). But, as we have already pointed out, the suffix *-in* does appear further to the south, it just takes on other shapes, apparently due to the influence of German dialects. Known differences between Polabian and Sorbian are too subtle to be found in a broad, pattern-seeking approach using modern place names. A major difficulty is that the Slavic names became blurred by German orthographic conventions as the area was appropriated by these non-Slavic populations starting around a millennium ago. An even greater difficulty is changing pronunciations, including many changes due to accommodation to German.

Instead, when we do find diagnostic traits they seem to reflect German dialect features rather than Slavic ones. For instance, an area quite closely corresponding to Berlin-Brandenburg stands out by exhibiting hydronyms containing the lexemes *Fließ*, *Pfuhl*, and *Luch*, as shown in figure 12.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

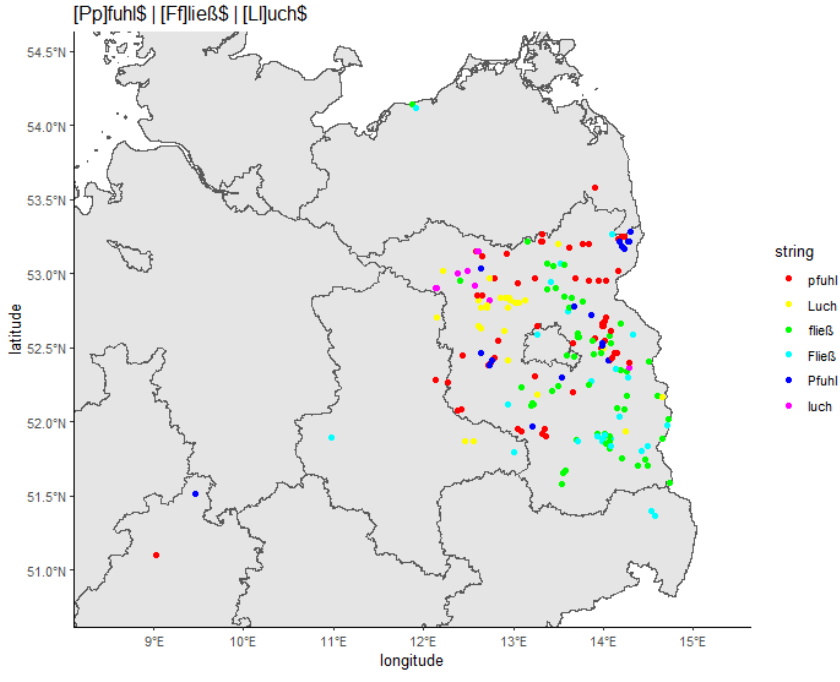

Figure 12: Distribution of the Lexemes *Pfuhl*, *Luch*, and *Fließ* in Hydronymic Compounds

Fließ means 'stream, river, canal'. It comes from Proto-Germanic *fleuta, according to Greule (2014, 150), who does not comment on the distribution. *Pfuhl* means 'pond, puddle'. It is related to Low German *Pool*, *Puhl*, *Paul*, *Puel*. In the 18th century larger bodies of water could be referred to as *Pfuhl* (cf. Greule 2014, 405). *Luch* means 'swamp'. Its origin is West Slavic, ultimately reflecting Proto-Slavic *lǫg 'meadow' (Debus & Schmitz 2002), and it apparently predates the split between Polabian and Sorbian. Thus, the distribution of these lexical items seem to mainly be characteristic of German dialect differences. To claim that their distributions somehow relate to a Slavic substrate influence would be speculative.

Discussion

Two case studies were presented. In the first one, on Xincan, we demonstrated how maps drawing upon GeoNames data inform proposals concerning place names that have been taken up again repeatedly since they were first proposed around half a century ago (Campbell 1978; Thompson 1970). Thompson proposed that the ending *–agua* in place names of Guatemala, Honduras, El Salvador, and Nicaragua could be connected to a specific ethnic group, and Campbell proposed that some of these names belong to the Xincan family of languages. In both cases the proposals are made in the absence of maps. When mapping the names discussed, Thompson's proposal is brought out more clearly, and the widely scattered distribution of names in *–agua* casts doubts on the viability of a hypothesis linking them to a certain language or language groups. Campbell's proposal is better supported, but this proposal also benefitted from mapping, which brought out the need for some corrections.

In the second case study, we exercised the GeoNames data in the context of a better-studied toponymic research area, namely that of Slavic place names in Germany. The package is only as good as the data for which it provides an interface, which consists of modern place names. This puts severe restrictions on a linguistic investigation where the shapes of toponyms as they first appear in attestations from the Middle Ages would be crucial. Nevertheless, even for a study area privileged by masses of literature the software is useful, especially as a companion for quickly producing distributional maps of morphological and lexical elements brought up in the literature.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

Conclusion

In this paper we have illustrated uses of the new R package 'toponym', which is designed to select, display, and manipulate data from GeoNames. The package is a research tool particularly useful in a pattern-seeking approach to toponymy while the complementary particularistic approach that concentrates on individual elements and their histories requires additional tools like etymological dictionaries and archives. Our case studies on Xincan toponyms and Slavic toponyms in Germany were intended not only to illustrate the utility but also the limitations of the GeoNames data.

Notes

[1] GNIS is hosted at https://www.usgs.gov/tools/geographic-names-information-system-gnis

[2] https://github.com/Sokiwi/ToponymPaper/

[3] https://www.geonames.org/datasources/

[4] http://download.geonames.org/export/dump/readme.txt

[5] There is in a sense one name 'too many' because *Río Sansur* is repeated in our output. We removed the westernmost occurrence from the data.

[6] The Motagua River runs just below 15° N.

[7] A more in-depth study of the earlier distribution of Xincan should also take into account records from the colonial period (cf. Sachse 2010, 42–47).

Acknowledgements

We thank Bernard Comrie for comments on the manuscript.

Funding Details

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy, Grant number EXC 2150 390870439.

References

Acheson, Elise, Stefano De Sabbata, and Ross S. Purves. 2017. "A Quantitative Analysis of Global Gazetteers: Patterns of Coverage for Common Feature Types". *Computers, Environment and Urban Systems* 64: 309–320.

Algeo, John and Katie Algeo. 2000. "Onomastics as an Interdisciplinary Study". *Names* 48, no. 3/4: 265–274.

Azaryahu, Maoz and Arnon Golan. 2001. "(Re-)naming the Landscape: The Formation of the Hebrew Map of Israel 1949–1960". *Journal of Historical Geography* 27, no. 2: 178–195.

Baca, Keith A. 2007. *Native American Place Names in Mississippi*. Jackson: University Press of Mississippi.

Bathe, Max, Reinhard E. Fischer, and Gerhard Schlimpert. 1971. "Die sorbisch-polabische Sprachgrenze und das Problem der Entnasalierung im 10. bis 12. Jahrhundert" [The Sorbian-Polabian Language Border and the Problem of Denasalization in the 10th to the 12th Centuries]. *Forschungen zur slawischen und deutschen Namenkunde* [Research on Slavic and German Onomastics] (= Deutsche Akademie der Wissenschaften. Veröffentlichungen des Instituts für Slawistik, 55), edited by Teodolius Witkowski, 32–39. Berlin: Akademie-Verlag.

Bily, Inge. 2002. "Zu einigen namenkundlichen Projekten an der Sächsischen Akademie der Wissenschaften zu Leipzig (1992–2003)" [On some Onomastic Projects in the Saxonian Academy of Sciences of Leipzig (1992–2003)]. *Onoma* 36: 7–24.

Bily, Inge. 2003. “Siedlungsnamen und slavische Besiedlung” [Settlement Names and Slavic Settlement]. *Słowiańska onomastyka* [Slavic Onomastics], Vol. 2, edited by Ewa Rzetelska-Feleszko, Aleksandra Cieślikowa, and Jerzy Duma, 21–22. Warszawa-Kraków: Towarzystwo Naukowe Warszawskie.

Bily, Inge. 2015. “Deutsch-slawischer Siedlungs- und Sprachkontakt im Gebiet zwischen Saale und Neiße – vorgestellt an ausgewählten Ortsnamen (Siedlungsnamen)” [German-Slavic Settlement and Language Contact in the Area between Saale and Neiße – Presented through Selected Place Names (Settlement Names)]. *Linguistica* 55, no. 1: 131–147.

Bright, William. 1998. *1500 California Place Names. Their Origin and Meaning. A Revised Version of 1000 California Places Names, by Erwin G. Gudde,* 3rd ed. Berkeley: University of California Press.

Campbell, Jon C. 1991. “Stream Generic Terms as Indicators of Historical Settlement Patterns”. *Names* 39, no. 4: 333–366.

Campbell, Lyle. 1978. “Quichean Prehistory: Linguistic Contributions”. *Papers in Mayan Linguistics*, edited by Nora C. England, 25–54. Columbia: Museum of Anthropology, University of Missouri.

Campbell, Lyle. 1998. “The Linguistic Prehistory of Guatemala”. *The Life of Language: Papers in Honor of William Bright*, edited by Jane Hill, P. J. Mistry, and Lyle Campbell, 183–192. Berlin/New York: Mouton de Gruyter.

Campbell, Lyle. 2013. *Historical Linguistics: An Introduction*, 3rd ed. Edinburgh: Edinburgh University Press.

Campbell, Lyle. 2022. “Las lenguas indígenas de El Salvador” [The Indigenous Languages of El Salvador]. Talk presented (online) at Tamachtiluyan Alan R. King, Escuela Virtual del Náhuat de El Salvador, October 21, 2022.

Chen, Xixi, Tao Hu, Fu Ren, Deng Chen, Lan Li, and Nan Gao. 2014. “Landscape analysis of geographical names in Hubei Province, China”. *Entropy* 16: 6313–6337. https://doi.org/10.3390/e16126313

Chevallier, Lennart, and Søren & Wichmann. 2024. “A Note on the ‘toponym’ R Package: A Practical Introduction”. *Names* 72, no. 3: 77-83.

Coates, Richard. 2012. “Onomastics”. *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle. Blackwell Publishing Ltd.

Cohen, Saul B. and Nurit Kliot. 1992. “Place-names in Israel's Ideological Struggle over the Administered Territories”. *Annals of the Association of American Geographers* 82, no. 4: 653–680.

Copley, Gordon. 1988. *Early Place-Names of the Anglian Regions of England*. BAR British Series, 185. Oxford: B.A.R.

Debus, Friedhelm and Antje Schmitz. 2002. “(Mikro-)Toponyme im slawisch-deutschen Kontaktgebiet Norddeutschlands” [(Micro) Toponyms in the Slavic-German Contact Area of Northern Germany]. *Onoma* 36: 51–70.

DeRosier, Jr., Arthur H. 1970. *The Removal of the Choctaw Indians*. New York: Harper & Row.

Eichler, Ernst. 1976. “Sprachkontakte im Lichte der Onomastik” [Language Contact in the Light of Onomastics]. *Onoma* 20: 128–141.

Eichler, Ernst. 1994–95. “Namenkundliche Forschungen in Leipzig 1990-1995” [Onomastic Research in Leipzig 1990–1995]. *Onoma* 32: 23–25.

Eichler, Ernst. 1998. “Die westlichste Peripherie des slavischen Sprachgebietes” [The Westernmost Periphery of the Slavic Language Area]. *Zeitschrift für Slavische Philologie* 57, no. 2: 269–280.

Eichler, Ernst and Hans Walther. 2001. *Historisches Ortsnamenbuch von Sachsen* [Historical Book of Place Names of Saxony]. Vol. 1–3. Berlin: Akademie Verlag.

Eichler, Ernst and Teodolius Witkowski. 1985a. “Namen der Stämme und Landschaften” [Names of the Tribes and Landscapes]. *Die Slawen in Deutschland. Geschichte und Kultur der slawischen Stämme westlich von Oder und Neiße vom 6. bis 12. Jahrhundert* [The Slavs in Germany. History and Culture of the Slavic Tribes West of Oder and Neiße from the 6th to the 12th Centuries], edited by Joachim Hermann, 10–14. Berlin: Akademie-Verlag.

Eichler, Ernst and Teodolius Witkowski. 1985b. “Sprachen und Dialekte” [Languages and Dialects]. *Die Slawen in Deutschland. Geschichte und Kultur der slawischen Stämme westlich von Oder und Neiße vom 6. bis 12. Jahrhundert* [The Slavs in Germany. History and Culture of the Slavic Tribes West of Oder and Neiße from the 6th to the 12th Centuries], edited by Joachim Hermann, 47–54. Berlin: Akademie-Verlag.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616

Endo, Mitsuaki. 2021. "Geographical Distribution of Certain Toponyms in the Samguk Sagi". *Anthropological Science* 129, no. 1: 35–44.

Fischer, Reinhard E. 1970. "Zur sorbisch-polabischen Sprachgrenze zwischen Elbe und Spree" [On the Sorbian-Polabian Language Border between Elbe and Spress]. *Beiträge zum Slawischen Onomastischen Atlas. Theodor Frings zum Gedächtnis* [Contributions to the Slavic Onomastic Atlas. In Remembrance of Theodor Frings] (= Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Phil.-hist. Kl. 61/2), edited by Rudolf Fischer and Ernst Eichler, 190–221. Berlin: Akademie-Verlag.

Förstemann, Ernst Wilhelm. 1863. *Die deutschen Ortsnamen* [The German Place Names]. Nordhausen: Ferd. Förstemann's Verlag.

Fuchs, Stephan. 2015. "An integrated approach to Germanic place names in the American Midwest". *The Professional Geographer* 67, no. 3: 330–341.

GeoNames. 2023. Accessed December 7, 2023. http://www.geonames.org/

Greule, Albrecht. 2014. *Deutsches Gewässernamenbuch. Etymologie der Gewässernamen und der dazugehörigen Gebiets-, Siedlungs- und Flurnamen* [Book of German Hydronyms. Etymologies of the Hydronyms and Related Names of Areas, Settlements, and Plots of Land]. Berlin: De Gruyter.

Gringmuth-Dallmer, Eike. 1990. "Deutsch und Wendisch - Groß und Klein. Zur Siedlungsgeschichtlichen Aussage von Ortsnamen mit unterscheidenden Zusätzen in der Mark Brandenburg" [*Deutsch* and *Wendisch* – Big and Small. On Information about Settlement History Contained in Place Names with Different Additions in Mark Brandenburg]. *Onomastica Slavogermanica XIX. Hans Walther zum 70. Geburtstag* [Onomastica Slavogermanica XIX. For Hans Walther on His 70th Birthday]. (= Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Philologisch-Historische Klasse, Band 73, Heft 2), edited by Ernst Eichler, 77–89. Berlin: Akademie-Verlag Berlin.

Hermann, Joachim. 1985. "Wohnsitze der slawischen Stämme und Stammesgruppen" [Dwelling Places of the Slavic Tribes and Tribal Groups]. *Die Slawen in Deutschland. Geschichte und Kultur der slawischen Stämme westlich von Oder und Neiße vom 6. bis 12. Jahrhundert* [The Slavs in Germany. History and Culture of the Slavic Tribes West of Oder and Neiße from the 6th to the 12th Centuries], edited by Joachim Hermann, 7–10. Berlin: Akademie-Verlag.

Hunn, Eugene. 1996. "Columbia Plateau Indian Place Names: What Can They Teach Us?" *Journal of Linguistic Anthropology* 6, no. 1: 3–26.

Lehmann, Friedrich. 1999. *Der Wandel der Ortsnamen in den ehemals deutsch besiedelten Gebieten der Tschechoslowakei. Gezeigt an über 300 Beispielen ausgewählter ehemaliger Landkreise* [Changes in Place Names in the Areas of Czechoslovakia Formerly Settled by Germans. Shown Through More Than 300 Examples of Selected Former Districts]. (= Scripta Slavica, Band 6). Marburg/Lahn: Biblion Verlag.

Perono Cacciafoco, Francesco and Francesco Cavallaro. 2023. *Place Names: Approaches and Perspectives in Toponymy and Toponomastics*. Cambridge: Cambridge University Press.

Petersen, Niels Matthias. 1833. "Bemærkninger om danske og norske stedsnavnes oprindelse og forklaring" [Remarks on the Origins of and Explanations for Danish and Norwegian Place Names]. *Nordisk tidsskrift for oldkyndighed* [Nordic Journal of the Study of Antiquity] 2.

Plate, Christa. 1998. "Die Besiedlungsgeschichte des Landes Ruppin" [The Settlement History of the Ruppin District]. *Brandenburgisches Namenbuch*. Teil 11: *Die Ortsnamen des Landes Ruppin* [The Book of Names of Brandenburg. Part 11: The Place Names of the Ruppin District], edited by Elżbieta Foster, 9–35. Weimar: Verlag Hermann Böhlaus Nachfolger.

R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Radding, Lisa and John Western. 2010. "What's in a Name? Linguistics, Geography, and Toponyms". *The Geographical Review* 100, no. 3: 394–412.

Read, William A. 1994. *Indian Place Names in Alabama*. Tuscaloosa and London: The University of Alabama Press.

Sachse, Frauke. 2010. *Reconstructive Description of Eighteenth-century Xinka Grammar*. Utrecht: LOT.

Tent, Jan. 2015. "Approaches to Research in Toponymy". *Names* 63, no. 2: 65–74.

TGN. 2023. "Getty Thesaurus of Geographic Names: About the TGN". Accessed November 7, 2023. http://www.getty.edu/research/tools/vocabularies/tgn/about.html

Thompson, J. Eric S. 1970. *Maya History and Religion*. Norman: University of Oklahoma Press.

DOI 10.5195/names.2025.2616 Vol. 73 No. 2, Summer 2025 ISSN: 0027-7738 (print) 1756-2279 (web)

Thorndahl, William. 1963. “Slawische Ortsnamen in Dänemark” [Slavic Place Names in Denmark]. *Slawische Namenforschung. Vorträge auf der II. Arbeitskonferenz der Onomastischen Kommission beim Internationalen Slawistenkomitee in Berlin vom 17. bis 20. Oktober 1961* [Slavic Onomastics. Presentations at the II. Working Conference of the Onomastic Commission of the International Committee of Slavists from the 17th to the 20th of October 1961], edited by Teodolius Witkowski, 56–70. Berlin: Akademie-Verlag.

Walther, Hans. 1995. “Namenforschung in der ehemaligen Deutschen Demokratischen Republik (DDR)” [Onomastics in the Former German Democratic Republic (GDR)]. *Namenforschung / Name Studies / Les noms propres. Ein internationales Handbuch zur Onomastik / An International Handbook of Onomastics / Manuel international d'onomastique. = Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science / Manuels de linguistique et des sciences de communication*. Vol 11.1, edited by Ernst Eichler, Gerold Hilty, Heinrich Löffler, Hugo Steger, and Ladislav Zgusta, 102–124. Berlin: Walter de Gruyter.

Notes on Contributors:

Søren Wichmann is a Postdoctoral Fellow at Kiel University (since 2021). He has previously held positions at University of Copenhagen, Leiden University, Kazan Federal University, and Max Planck Institute for Evolutionary Anthropology. He specializes in historical linguistics, descriptive linguistics, language typology, quantitative methods, and Mesoamerican languages and writing systems.

Lennart Chevallier holds an MA in Language, Variation, and Political Science at Kiel University. He is currently a PhD candidate and is a part of the “DiaSAL” project at Kiel University.

Correspondence to: Søren Wichmann, ROOTS, CAU, Leibnizstraße 3, 24118 Kiel, Germany; Email: wichmannsoeren@gmail.com

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 73 No. 2, Summer 2025 DOI 10.5195/names.2025.2616