Extracting Toponyms from OpenStreetMap: A Cross-Linguistic Perspective

Francesco-Alessio Ursini[1,†], Giuseppe Samo[2,3,*,†]

[1]Central China Normal University, School of Chinese Language and Literature, 52, Dailyou Road, Wuhan, 625762, China

[2]University of Geneva, Department of Linguistics, Rue de Candolle 2, Geneva, 1205, Switzerland

[3]Beijing Language and Culture University, 15, Xue Yuan Road, Beijing, 100083, China

Abstract

In this paper we discuss three studies in which we performed toponym extraction from OpenStreetMap. The studies operated at three different levels of geographical resolution: city level (Macao), national level (Italy and its regions), and province level (Geneva and its district). We present a single algorithm that we used in each study to extract toponyms from the text database associated to the corresponding OSM map. For each study, we provide a summary of the results, some observations on the language-specific methodological and theoretical aspects, and language-general/cross-linguistic considerations. We conclude by analyzing the reliability of OSM for toponym extraction and linguistic theory.

Keywords

OpenStreetMap, Toponym recognition, Toponym extraction, Cross-linguistic analysis

1. Introduction

OpenStreetMap (henceforth: OSM; https://openstreetmap.com) is an on-line platform that offers "a free, editable map of the world", since its inception in 2004 [1, 2, 3]. OSM is a clear-cut case of a platform implementing a *Volunteered Geographic Information* philosophy (henceforth: VGI; [4, 5, 6, 7]). Registered contributors can insert and edit information via their knowledge of locations. Contributions center on the geographical objects shaping maps: "nodes" representing locations, "ways" representing connections among locations, and "relations" between nodes and/or ways [8, 9]. Each object has "tags", labels indexing attributes ("keys") and values associated to locations (e.g. coordinates, altitude, shape, type of location). Hence, OSM chiefly offers information about "places": locations in which humans perform activities and to which they develop attachment relations, possibly via the names bestowed to these places [10, 11].

Studies analyzing OSM-based data in GIS cover an expansive domain of topics (e.g. [12]). Several works investigate the accuracy of OSM maps and data when compared to official

GeoExT 2023: First International Workshop on Geographic Information Extraction from Texts at ECIR 2023, April 2, 2023, Dublin, Ireland

*Corresponding author.

†These authors contributed equally.

ursini@ccnu.edu.cn (F. Ursini); giuseppe.samo@unige.ch/samo@blcu.edu.cn (G. Samo)

https://github.com/samo-g (G. Samo)

0000-0001-7042-3576 (F. Ursini); 0000-0003-3449-8006 (G. Samo)

gazetteers (e.g. Ordnance surveys) and other *Authoritative Geographic Information* sources (henceforth: AGI, [13, 14, 15, 16, 17]). Many AGI sources have become open access (e.g. official gazetteers, Google Maps) and thus freely accessible to the public. Hence, contributors can combine these AGI-based data with personal volunteered knowledge, thus contributing to large data imports [18]. Many works have also studied how OSM can provide real-time and long-term information regarding complex situations and risks affecting places (e.g. natural disasters: [19, 20, 21]; epidemic diffusion: [22, 23]). OSM maps can therefore offer highly detailed, synchronous geo-located information, often thanks to local contributors' direct knowledge. However, the soundness of this information usually correlates with contributors' formal education, motivation and commitment to rigorous, professional-like data insertion [24, 25, 26, 27, 28].

Studies using OSM for the retrieval and analysis of toponyms, i.e. names for places, seem to represent an emerging field of study (e.g. [29, 30, 31, 32]). In OSM toponyms act as labels for tags; they are linked to content-rich descriptions of objects, and thus to the places that these objects represent [33, 34]. The increasing integration of AGI and VGI has contributed to the growth of toponyms mapping in OSM. For instance, the Parisian toponym database has experienced a fast expansion due to contributors' access of public gazetteers, combined with their grassroots knowledge [35]. The Jerusalem toponym database seems heavily biased towards Hebrew toponyms, but the insertion of Palestinian Arabic toponyms is currently gaining heavy momentum, too [36]. Similar patterns of increasing toponym coverage are attested across regions and cultures (e.g. China: [37]; Italy: [15]; Kenya: [38]). OSM and its users benefit from this dramatic increase of coverage. However, such a situation raises questions about this empirical growth and its theoretical consequences for geographic and linguistic disciplines.

The goal of this paper is to answer two questions that arise from the evolving status of OSM as a toponym database. The first question is how reliable can be toponym extraction via OSM, so that it can feed linguistic and toponomastic (i.e. interdisciplinary) analysis. We thus present three studies in which we implemented a single extraction algorithm for OSM data and AGI sources. We then discuss the quantitative and qualitative aspects of these results. The second question is how accurate can be OSM-based data when used for cross-linguistic analysis, i.e. analysis involving data from several languages. We show that by using OSM, one can operate at fine-grained levels of resolution irrespective the scale of analysis, and provide detailed linguistic analyses of toponyms across languages. Section 2 presents the general methodology; Section 3 the specific studies and results; Section 4 offers a discussion and conclusions.

2. Methodology

We used one methodological approach across our three studies; we discuss language-specific adjustments in section 3. The first study features the city of Macao, one of the two China's special administrative regions. The second study features Italy as a national territory, though it also involves the analysis of toponyms at a regional resolution. The third study features the district/province of Geneva, Switzerland. The respective data sets were used for the creation of language-specific studies focusing on various unaddressed linguistic problems (i.e. [39, 40, 41]). Here we report a meta-analysis of the studies with a cross-linguistic import. Given the linguistic focus of these studies, we did not discuss fine-grained methodological problems, as they were not

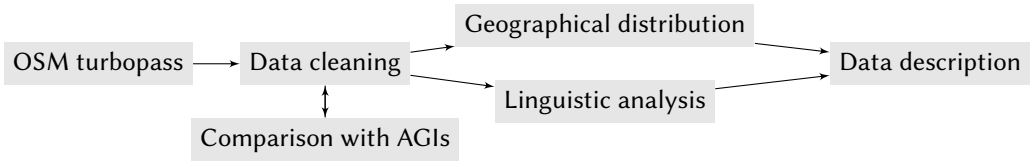

Figure 1: Flowchart underpinning the methodology used in each study.

crucial to the study-specific goals. Hence, the cross-linguistic comparison was still outstanding.

The methodology worked as follows (cf. also Fig. 1 for the flow-chart). We accessed to OSM data through the the platform we used the platform overpass-turbo (https://overpass-turbo.eu/), sizing our search in the relevant geographical areas. The output .csv file easily supports statistical data analysis, visual data representation and linguistic categorization. We aimed to achieve a form of triangulation, i.e. to verify that the same data set via two partially different methods [42, 43]. We thus compared the OSM data with data extracted from AGI sources specific to the regions under study. After obtaining the toponyms data, we turned to the linguistic analysis of these data together with the geographical distribution. We performed frequency and geographical distributions analyses so that toponyms could be organised according to the types and research questions specific to each study. We discuss the questions underpinning our study and the problems that emerged relative to the questions in the next Section.

3. The studies

3.1. First Study: Toponyms in Macao

In the first study [39], we investigated the grammatical and lexical properties of Macanese toponyms. Macao is a city and a special administrative region in the Pearl River Delta, South-East China [44]. Cantonese is the most commonly spoken language (90% of the population), and has official status along Portuguese, which is however slowly disappearing (2% of the population: [45]). Toponyms are reported in gazetteers and atlases in the Portuguese and Chinese written systems [46]. Macanese Portuguese is near-identical to European Portuguese; thus, no spelling differences exist in Macanese toponyms. Cantonese toponyms are written in the Chinese simplified characters system, and are thus intelligible to speakers of Cantonese and other Sinitic languages (e.g. Mandarin, Hakka). Our goal was to analyse the internal order of toponym constituents in both Macanese languages (i.e. their grammatical properties), and analyse how toponyms could potentially classify places via their lexical content/meaning.

The details of our methodology for this study were as follows. In the first step, one researcher extracted Portuguese and Chinese OSM toponyms. In the second step, two researchers compared these data with an official gazetteer in CD-ROM form, as an AGI source [47]. The researchers obtained two lists of 1394 toponyms, comparing the tokens on a case-by-case basis. We performed an analysis of the Jaccard Index of similarity (from 0 to 1: the closer to 1, the more similar two populations, [48]) between the two lists, and obtained a 0.989 as a result. Qualitatively, the only difference was that Portuguese toponyms featured minor spelling variants in OSM (e.g. accent omission: *Rua de Santo Antonio* instead of *Rua de Santo António* 'Saint Anthony street'). Aside

this minor divergence, we could confirm that the OSM data were as accurate as those found in the CD-ROM. We thus conjecture that OSM contributors have not inserted information about places not attested in official maps (e.g. shops, cafes, and other smaller places).

The analysis overall showed that toponym extraction via OSM proved as reliable as extraction via the CD-ROM gazetteer: our first question finds a positive answer. Furthermore, the OSM data lent themselves to a direct linguistic comparison of Chinese and Portuguese toponyms: the dual set of tokens confirmed that all places on the map(s) included two toponyms. For instance, the two toponyms 龍鬚街 *lung4 sou1 gaai1* 'Dragon Beard Street' and *Rua Central* 'Central Street' reported the same geo-coordinates in their database entry. This fact showed that they were the respective Chinese and Portuguese names for this place. Our second question also finds a positive answer. Overall, our methodology successfully allowed us to extract the Portuguese and Chinese toponyms from OSM, obtain an exhaustive number of tokens for both languages. It also allowed us to perform a (cross-)linguistic analysis and comparison of these systems and their origins, therefore shedding light on the linguistic properties of these toponyms.

3.2. Second Study: Toponyms in Italy

In the second study [40], we investigated the distribution of toponyms of dialectal origin in Italy. Italian toponyms often find their roots in the languages of the pre-Italic populations that once inhabited Italy, but also in the local dialects spoken across the country [49, 50]. Since dialectal toponyms are reported via standard Italian spelling in gazetteers and other AGI sources, we investigated the possibility that dialectal toponyms correlate in geographical distribution with their dialects. We then used the output .csv file to create a database of toponyms organized according to their distribution in each of the 20 Italian administrative regions. We could thus compare toponyms distribution with the dialect(s) spoken in each region.

We obtained 452,538 toponyms, with distribution among regions being quite uneven. Regions with higher numbers of cities and urban centers offered the highest number of toponyms. For instance, Milan and its region Lombardy covered 11.36% of the total; the capital city of Rome and its region Lazio 8.11% of the total toponyms. Small regions such as Abruzzo, in central Italy, only offered 2.91% of the total. Nevertheless, all regions and major dialects offered evidence for toponyms of local/dialectal origins. This is the case because for toponyms displaying clear dialectal origins, we could observe that their distribution was often limited to the regions in which a toponym was found. For instance, toponyms for streets in Venice and the surrounding Veneto region may include the generic (i.e. classificatory) term *calle*, literally 'narrow alley'. Outside this region, we only found sporadic cases of toponyms including *calle*; this is strong evidence that these toponyms have Venetian (i.e. dialectal) origins.

For the goals of our paper, the study offers two answers. First, OSM critically offers a higher number of toponyms than some AGI sources. We compared this result with a previous study in which we extracted toponyms from the YellowPages online directory (https://paginegialle.it; [51]). In this latter study, we extracted toponyms from province-based directories, with "province" being the immediate administrative unit below regions. We obtained 213,218 toponyms, i.e. less than half of the toponyms extracted via OSM. One possible explanation is that the YellowPages directory includes directories for minor urban centers, villages and hamlets. However, these directories tend to offer lower- resolution maps than those for major urban

centers (e.g. 1:5000 against 1:3000). Non-urban places are also absent, and one must use dedicated gazetteers for retrieving those toponyms (e.g. plates for mountainous ranges: [50]). We also computed a Jaccard index on these lists, and the result was clear (0.088): the YellowPages directory data represent a sub-set of the OSM data. OSM thus now provides a more reliable and a faster source for toponym extraction, at least for the regional and national scale level of Italy.

Second, OSM did not offer cues on the dialectal origins of toponyms. We had to analyse the lexical properties of toponyms (e.g. their etymologies and senses), and their geo-linguistic distribution and correlation with dialects. The higher number of toponyms also entailed that we could access toponyms not reported in the YellowPages gazetteers. For instance, the Sicilian term *ronco* designates a type of narrow alley mostly found in the city of Syracuse and in other cities from this island. In OSM, we could find several instances of toponyms including this term. In the YellowPages directory these terms were missing, possibly due to these alleys being too small to appear in the gazetteers' 1:3000 maps, compared to OSM's average scale of 1:1000. Hence, OSM data may require further linguistic analysis when one attempts comparisons across languages, but these data may be quantitatively superior than in AGI sources.

3.3. Third Study: Toponyms in Geneva

In the third study [41], we investigated the lexicographic properties of toponyms in Geneva and its surrounding canton (i.e. district). Geneva is a Francophonic enclave in Switzerland and a global hub for diplomatic and economic communities (see [52, 53] for some sociological studies). Geneva's district represents a departure point for the lush mountainous environments and natural attractions for which Switzerland attracts millions of tourists and acts as a target for "scientific tourism" [54]. We therefore studied the geo-linguistic distribution of toponyms referring to typical urban and rural places and their linguistic properties. For this study, we used two sources for database creation. One is the platform overpass-turbo; the other is the dedicated repository *Noms géographiques du canton de Genève* (https://noms-geographiques.app.ge.ch/), an official on-line gazetteer for the Geneva canton.

We found 3843 toponyms in the official repository, 3713 in OSM, thus detecting a small asymmetry. We analyzed the two lists to verify that the toponyms matched (Jaccard index: 0.680). The city of Geneva covered most of the attested toponyms (i.e. 918 tokens in OSM (24.72% of the total); 927 tokens (24.12%) in the dedicated repository). The Geneva district covered the remainder of toponyms, with the municipalities of Meyrin (5.47% OSM; 4.81% repository), and Lancy (4.23% OSM; 4.24% repository). Furthermore, a lexicographic analysis suggested that Geneva and nearby cities mostly featured toponyms for urban places (e.g. *rue* 'urban street' and *avenue* 'broad road in a urban agglomeration'). Conversely, rural zones made up the rest of the district's toponyms (e.g. *route* 'road' and *chemin* 'path'). The sharp distinction between urban and rural territories was mostly mirrored in the toponyms lexicon for this Canton.

We obtained two key answers for our target questions. First, OSM and repository seem to minimally diverge in their coverage. Since the repository is an open access platform, we conjecture that OSM contributors may have uploaded the data semi-manually from the repository, though this process is still undergoing. Second, and likely a consequence of the first result, the distribution of toponyms correlated with urban centers. Outside urban centers, paths but also touristic attractions (i.e. mountainous paths or *chemins* in French) are the chief objects

also including information about toponyms. Genevian OSM data therefore seem highly reliable though less so than the repository as its AGI counterpart, and amenable to linguistic analysis. They may also provide a relatively clear geo-linguistic picture on toponyms types' distribution.

4. Discussion

We propose two general results as language-general answers to our questions.

The first result is that toponym extraction from OSM seems now highly reliable, modulo the use of dedicated algorithms and databases providing geographically-bound toponyms lists (e.g. Macanese toponyms). Reliability also stems from the fact that OSM as a VGI source can provide data that may be quantitatively and qualitatively superior to AGI sources (cf. [7, 10, 17, 24]). Contributors, after all, can import toponyms from official gazetteers when available (cf. [25, 33, 35, 38]). Arguably, we tested three Romance languages (Portuguese, French, and Italian) that involve minor differences in spelling standards (e.g. use of accents and diacritics). Only in the case of Portuguese did we find minor spelling omissions (i.e. accents) in OSM, which nevertheless did not affect our results. Furthermore, the retrieval of Chinese toponyms in the Macao study also did not provide any challenges to the algorithm. Overall, the algorithm successfully retrieved all the toponyms from each database and writing system.

The second result is that OSM-based data now provide accurate information on which to develop a cross-linguistic analysis that builds on language-specific results. From the first study, we know that Macanese places have dual (i.e. Portuguese, Chinese) toponyms. From the second study, we know that Italian toponyms often have dialectal, geographically specific roots. From the third study, we know that Geneva and district include a high number of toponyms for urban places. Hence, each study provides a language-specific set of results. The three studies also show that OSM and other AGI sources include similar data-sets, as we summarise in Table 1:

Study	OSM toponyms	AGI toponyms	Jaccard Index
Macao	1,394	1,394	0.989
Geneva	3,713	3,843	0.680
Italy	452,538	213,218	0.088

Table 1
Investigated areas, number of OSM toponyms, number of AGI toponyms and Jaccard Index.

We conjecture that these results may be correlated with the size of the places and corresponding toponym lists that we targeted in our studies (cf. also [55] for discussion). The data from the first study (the city of Macao) suggest that AGI and VGI sources can be identically detailed on places providing smaller, well-documented toponym lists. The data from the second study (province of Geneva) suggest that AGI sources may still offer more detailed pictures. However, the data from the third study (Italy and its regions) suggest that the opposite trend may become the norm in the future. Thus, OSM as a VGI source may slowly turn into a more accurate and thorough source for toponyms, since it can integrate data from several AGI's into a single map.

From these results, certain cross-linguistic generalizations with a geo-linguistic import logically emerge. First, toponyms often appear as compound nouns, in which a classifier or "generic term" may either precede or follow a byname or "specific term". The analysis of this

category often found in Anglophonic toponomastics (e.g. [56, 57]) thus seems to have cross-linguistic import. Second, access to the type of high-resolution data that OSM can improve the quality of this interdisciplinary analysis. Overall, OSM may provide direct access to toponyms that may only be found in e.g. national/regional-, provincial/district- and city-based AGI gazetteers (e.g. [30]; [33]–[39]). Hence, it allows researchers to address cross- and geo-linguistic questions via one data extraction methodology, at least in our case.

5. Conclusion

This study has provided recent evidence that OSM can offer highly reliable data regarding toponyms. However, this reliability hinges on the regions of interest and contributors' access to AGI data and their nuanced knowledge of local toponyms. Hence, OSM data can now support single- and cross-linguistic generalizations: the meta-analysis of our three OSM-based studies supports this conclusion. This result entails that "platial" (i.e. place-based, [58]) research in linguistics can find a veritable methodological ally in OSM. Furthermore, the study has also shown that OSM can potentially provide a wealth of toponymic data: our algorithm extracted toponyms irrespective of the writing system in which toponyms were reported. We acknowledge that the choice of regions for such studies can heavily influence reliability. For regions on which OSM data may still not scale up to AGI sources, perhaps we may not obtain similar results. We leave this and other challenges for future studies.

References

[1] K. Curran, J. Crumlish, G. Fisher, Openstreetmap, International Journal of Interactive Communication Systems and Technologies (IJICST) 2 (2012) 69–78. doi:10.4018/ijicst.2012010105.

[2] K. Curran, J. Crumlish, G. Fisher, Openstreetmap, in: Geographic Information Systems: Concepts, Methodologies, Tools, and Applications, IGI Global, 2013, pp. 540–549. doi:10.4018/978-1-4666-2038-4.ch033.

[3] C. Keßler, Openstreetmap, Encyclopedia of GIS (2017) 1493–1498. doi:10.1007/978-3-319-17885-1_1654.

[4] M. F. Goodchild, Citizens as sensors: the world of volunteered geography, GeoJournal 69 (2007) 211–221. doi:http://dx.doi.org/10.1007/s10708-007-9111-y.

[5] C. Keßler, K. Janowicz, M. Bishr, An agenda for the next generation gazetteer: Geographic information contribution and retrieval, in: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in Geographic Information Systems, 2009, pp. 91–100.

[6] D. Sui, M. Goodchild, The convergence of GIS and social media: challenges for GIScience, International journal of geographical information science 25 (2011) 1737–1748. doi:10.1080/13658816.2011.604636.

[7] V. Antoniou, A. Skopeliti, The impact of the contribution microenvironment on data quality: the case of OSM, Mapping and the Citizen Sensor (2017) 165–196.

[8] A. Rajšp, M. Hericko, I. Fister Jr, Preprocessing of roads in OpenStreetMap based geographic

data on a property graph, in: Central European Conference on Information and Intelligent Systems, Faculty of Organization and Informatics Varazdin, 2021, pp. 193–199.
[9] J. M. Almendros-Jiménez, A. Becerra-Terón, M. G. Merayo, M. Núñez, Metamorphic testing of OpenStreetMap, Information and Software Technology 138 (2021) 106631. doi:138.106631.10.1016/j.infsof.2021.1066319.
[10] T. Cresswell, Place: an introduction, John Wiley & Sons, 2014.
[11] J. Malpas, Place and experience: A philosophical topography, Routledge, 2018.
[12] J. J. Arsanjani, A. Zipf, P. Mooney, M. Helbich, OpenStreetMap in GIScience, Lecture notes in geoinformation and cartography (2015) 324.
[13] E. Bortolini, S. P. Camboim, Mapeamento colaborativo de favelas com a plataforma openstreetmap collaborative slum mapping with Openstreetmap, Mapeamento participativo: tecnologia e cidadania (2019).
[14] J. Fize, L. Moncla, B. Martins, Deep learning for toponym resolution: Geocoding based on pairs of toponyms, ISPRS International Journal of Geo-Information 10 (2021) 818. doi:10.818.10.3390/ijgi10120818.
[15] G. Salvucci, L. Salvati, Official statistics, building censuses, and OpenStreetMap completeness in Italy, ISPRS International Journal of Geo-Information 11 (2022) 29. doi:10.3390/ijgi11010029.
[16] S. Garba, H. Musa, S. Bala, N. Hafiz, M. İya, H. Náiya, M. Mustapha, G. Sule, Quality Analysis of OpenStreetMap and Digital Elevation Data Based North-Western, FIG Congress 2022 Volunteering for the future - Geospatial excellence for a better living Warsaw, Poland, 11–15 September 2022 (2022).
[17] T. Holthaus, A. Thiemermann, Identifikation deutscher Straßenentwurfsklassen im Straßennetz von OpenStreetMap, Road Network (2022). doi:10.14627/537728010.
[18] R. Witt, L. Loos, A. Zipf, Analysing the Impact of Large Data Imports in OpenStreetMap, ISPRS International Journal of Geo-Information 10 (2021) 528.
[19] R. Hecht, C. Kunze, S. Hahmann, Measuring completeness of building footprints in OpenStreetMap over space and time, ISPRS International Journal of Geo-Information 2 (2013) 1066–1091. doi:https://doi.org/10.3390/ijgi2041066.
[20] M. Cerri, M. Steinhausen, H. Kreibich, K. Schröter, Are OpenStreetMap building data useful for flood vulnerability modelling?, Natural Hazards and Earth System Sciences 21 (2021) 643–662. doi:21.10.5194/nhess-21-643-202.
[21] T. Seto, Development of OpenStreetMap Data in Japan, in: Ubiquitous Mapping, Springer, 2022, pp. 113–126.
[22] P. Mooney, L. Juhász, Mapping COVID-19: How web-based maps contribute to the infodemic, Dialogues in Human Geography 10 (2020) 265–270.
[23] P. Mooney, A. Y. Grinberger, M. Minghini, S. Coetzee, L. Juhasz, G. Yeboah, OpenStreetMap data use cases during the early months of the COVID-19 pandemic, COVID-19 Pandemic, Geospatial Information, and Community Resilience: Global Applications and Lessons (2021) 171–186.
[24] R. Jaljolie, T. Dror, D. N. Siriba, S. Dalyot, Evaluating current ethical values of OpenStreetMap using value sensitive design, Geo-spatial Information Science (2022) 1–17. doi:10.1080/10095020.2022.2087048.
[25] M. Mayer, D. W. Heck, F.-B. Mocnik, Using OpenStreetMap as a Data Source in Psychology

and the Social Sciences, PsyArXiv (2022). doi:10.31234/osf.io/h3npa.
[26] K. Wu, Z. Xie, M. Hu, An unsupervised framework for extracting multilane roads from OpenStreetMap, International Journal of Geographical Information Science 36 (2022) 2322–2344. doi:10.1080/13658816.2022.2107208.
[27] J. V. M. Bravo, C. R. Sluter, Crowdsourcing Map-Using and Map-Generating Tasks into OpenStreetMap, The Professional Geographer (2022) 1–15. doi:10.1080/00330124.2022.2094424.
[28] T. Novack, L. Vorbeck, A. Zipf, An investigation of the temporality of OpenStreetMap data contribution activities, Geo-spatial Information Science (2022) 1–17. doi:10.1080/10095020.2022.2124127.
[29] T. Schäfer, B. Kieslinger, Supporting emerging forms of citizen science: A plea for diversity, creativity and social innovation, Journal of Science Communication 15 (2016) Y02.
[30] A. P. Perdana, F. O. Ostermann, A citizen science approach for collecting toponyms, ISPRS international journal of geo-information 7 (2018) 222. doi:7.22.10.3390/ijgi7060222.
[31] M. Kaisar Ahmed, Converting OpenStreetMap (OSM) Data to Functional Road Networks for Downstream Applications, arXiv e-prints (2022) arXiv–2211. doi:10.48550/arXiv.2211.12996.
[32] A. A. Machado, E. N. N. Elias, L. S. L. Silva, S. P. Camboim, M. A. R. Schmidt, Informação geográfica voluntária: o potencial das ferramentas colaborativas para a aquisição de nomes geográficos, Education 2019 (2017).
[33] M. M. Hall, C. B. Jones, Generating geographical location descriptions with spatial templates: a salient toponym driven approach, International Journal of Geographical Information Science 36 (2022) 55–85. doi:10.1080/13658816.2021.1913498.
[34] S. Ahmadian, P. Pahlavani, Semantic integration of OpenStreetMap and CityGML with formal concept analysis, Transactions in GIS (2022). doi:10.1111/tgis.13006.
[35] V. Antoniou, G. Touya, A.-M. Raimond, Quality analysis of the Parisian OSM toponyms evolution, 2016. doi:10.5334/bax.h.
[36] V. Carraro, Naming Jerusalem on OpenStreetMap, Jerusalem Online: Critical Cartography for the Digital Age (2021) 87–109. doi:10.1007/978-981-16-3314-0_5.
[37] S. Qian, M. Kang, M. Wang, An analysis of spatial patterns of toponyms in Guangdong, China, Journal of cultural geography 33 (2016) 161–180. doi:https://doi.org/10.1080/08873.631.2016.1138795.
[38] N. Daniel, G. Mátyás, Citizen science characterization of meanings of toponyms of Kenya: a shared heritage, GeoJournal (2022) 1–22. doi:https://doi.org/10.1007/s10708-022-10640-5.
[39] Q. Xie, F.-A. Ursini, G. Samo, Urbanonyms in Macau, Names 71(1) (2023) 29–43.
[40] G. Samo, F.-A. Ursini, Geographical Maps meet Place Names where Languages meets Dialects: The Case of Italian (under review).
[41] G. Samo, F.-A. Ursini, Dictionnaire et Atlas: propriétés lexicales et sémantiques des urbanonymes en français (submitted).
[42] P. Rothbauer, Triangulation, The SAGE encyclopedia of qualitative research methods 1 (2008) 892–894.
[43] J. Damico, J. Tetnowski, Triangulation, in: C. Forsyth, H. Copes (Eds.), Encyclopedia of Social Deviances, Sage Publications, Riverside, Ca, 2014, pp. 709–721.

[44] H. Yee, The theory and practice of one country, two systems in Macau, China's Macao Transformed: Challenge and Development in the 21st Century. City University of Hong Kong Press, Hong Kong (2014) 3–20.
[45] A. J. Moody, Macau's Languages in Society and Education: Planning in a Multilingual Ecology, volume 39, Springer Nature, 2021.
[46] W. Botha, A. Moody, English in Macau, The Handbook of Asian Englishes (2020) 529–546.
[47] Cartography and Cadastre Bureau of Macau SAR, 澳門特別行政區數碼化地圖唯讀光碟A類/CD-ROM de Carta-base (Tipo A) da Região Administrativa Especial de Macau. [CD-ROM of the paper versión (type A) of the Special Administrative Region of Macau], 澳門特別行政區政府地圖繪製暨地籍/Macau: Direcção dos Serviços de Cartografia e Cadastro. (2021).
[48] P. Jaccard, Étude comparative de la distribution florale dans une portion des alpes et des jura, Bulletin de la Société vaudoise des sciences naturelles 37 (1901) 547–579.
[49] L. Cassi, P. Marcaccini, Toponomastica, beni culturali e ambientali. Gli indicatori geografici per un loro censimento, Collection Memorie della Società Geografica Italiana 55 (1998) 655–1097.
[50] C. Andrea, Lineamenti di storia della cartografia italiana, 2013.
[51] F.-A. Ursini, G. Samo, Names for urban places and conceptual taxonomies: the view from Italian, Spatial Cognition & Computation 22 (2022) 264–292. doi:10.1080/13875868.2021.1954186.
[52] S. Cattacin, F. Kettenacker, Genève n'existe pas. Pas encore? Essai sociologique sur les rapports entre l'organisation urbaine, les liens sociaux et l'identité de la ville de Genève, Genève à l'épreuve de la durabilité (2011) 29–36.
[53] F. Gamba, S. Cattacin, Urbans rituals as spaces of memory and belonging: A Geneva case study, City, culture and society 24 (2021) 100385.
[54] L. Molokáčová, Š. Molokáč, Scientific tourism–tourism in science or science in tourism, Acta Geoturistica 2 (2011) 41–45.
[55] R. Westerholt, The Analysis of Spatially Superimposed and Heterogeneous Random Variables, Ph.D. thesis, University of Heidelberg, 2019.
[56] D. Blair, J. Tent, Feature terms for Australian toponymy, ANPS Technical Paper (2015).
[57] D. Blair, J. Tent, A revised typology of Place Names, Names 69 (2021) 1–15.
[58] T. Tenbrink, The language of PLACE: Towards an agenda for linguistic platial cognition research, in: Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL-19), 2020, pp. 5–12. doi:10.5281/zenodo.3628849.