Putting Der Brenner on the map

Adrien Barbaresi

Putting *Der Brenner* on the map

Adrien Barbaresi

ÖAW-AC
Sonnenfelsgasse 19
1010 Wien Austria
`adrien.barbaresi@oeaw.ac.at`

Abstract

This article deals with the detection and projection of spatial patterns in the periodical *Der Brenner*. This case study highlights the interest of combining philological knowledge, geographic information retrieval and visualization. While the first prompts for further refinements, the latter two inform the interpretation of the text. Two different filter settings as well as maps focusing on the 1910s and 1920s are compared. Issues related to historical texts, data curation and preparation, processing of linguistic corpora and geocoding are presented and discussed.

1   Introduction

Although spatial studies have been a trend in the humanities since the 1970s, the interest for mapping techniques and geographic visualization is not self-explaining in a digital, globalized age where the emphasis lies on information, its standardization and its transmission. Space in an actual physical sense may even be considered to shift out of focus: "Globalization can be seen as an extension of modernity, due to its convergence with the Cartesian motif of eliminating the body as a part of human self-reference. Modernity and globalization thus imply a tendency of making us independent from the dimension of space." (Gumbrecht, 2009) That being said, despite of increasing independence from physical space, it seems to become paradoxically significant (*ibid.*), not least because of its (properly human) anthropological and linguistic dimension which exceeds a mere deictic frame.

In digital text corpora and especially literary works, space can be analyzed as pertaining to patterns to be found across works: "Large source bases are likely to contain 'meaningful patterns,' and the ability to examine them, that is, to gain a bird's-eye view with the assistance of a computer, is tantamount to having a team of readers, even many teams of readers, at one's disposal." (Wrisley, 2017) Looking for patterns is widely considered to be a task where distant reading and computer-based studies can be fruitful. In linguistics, the most common criteria of analysis are frequencies, but even in literary studies, patterns can be seen as "the strongest point of intersection between the computational strictures of text analysis and the open ended landscape of interpretative literary studies." (Ramsay, 2005)

The following approach deals marginally with frequency and more with a research object sometimes neglected by corpus linguistics, proper names and especially placenames. Looking for spatial patterns and displaying them can be assimilated to the recently coined concept of geocriticism, in the sense that it is geo-centered and not character-centered or subjective framework. Under this assumption, the emergence of geographical patterns in texts is explicitly fostered: "Geocriticism is a geo-centered rather than an ego-centered approach; that is, the analysis focuses on global spatial representations rather than on individual ones (a given traveler's, for example)." (Westphal, 2014)

This article documents distant, superficial reading experiments dealing with the detection and projection of spatial patterns in a historical text collection. As such it relates to a "post-disciplinary"

research field which can be called digital literary cartography (Cooper, 2017). The resulting outline is a global depiction which only marginally tackles the question of the semantic value of places in literary works[1] to offer instead a series of accesses and entry points (Barbaresi, 2018c). First, existing methods are described, then the issues in the context of *Der Brenner* as a text collection are introduced. Finally, results are presented and commented.

2 Trends in geocoding and visualization

Among the tendencies in geographic information retrieval and geocoding (Melo & Martins, 2017), the extraction and normalization of named places, itineraries, or qualitative spatial relations, as well as the extraction of locative expressions are particularly relevant to study text collections. The particular task of finding placenames in texts (geoparsing or placenames extraction) involves first the detection of words and phrases that may potentially be proper nouns and second their classification as geographic references (Nouvel, Ehrmann, & Rosset, 2015). After the identification of toponyms, a further step (geocoding or toponym resolution) resides in disambiguating and adding geographical coordinates to a place name. Geocoding mostly relies on gazetteers, i.e. geospatial dictionaries of geographic names, mostly names, locations, and metadata such as typological information, variants or dates (Hill, 2000). In this respect, knowledge-based methods using additional fine-grained registers have already been used with encouraging results. Especially for historical corpora, researchers face a lack of general-purpose tooling. In order to produce cartographic visualizations, both the capacity to adapt to different contexts (Alex, Byrne, Grover, & Tobin, 2015) and the necessity to complement existing resources with a precise historical gazetteer (Borin, Dannélls, & Olsson, 2014) have been highlighted. Such historical gazetteers exist, but their development is challenging (Southall, Mostern, & Berman, 2011) even for texts as late as 20th century Europe (Plini, Di Franco, & Salvatori, 2016).

3 Putting *Der Brenner* on the dissecting table

3.1 What is at stake?

Der Brenner is a periodical issued fortnightly under the direction of Ludwig von Ficker from 1910 until 1954, including several plurennial interruptions. The large majority of issues have been released between its foundation and the early 1930s. It is clearly stated objective at the beginning is to bring culture, literature and fine arts into focus. It soon integrates social criticism, especially until the 1st World War, features expressionism as a common denominator to a majority of texts until the 1920s, before turning more and more to philosophical and theological considerations. The journal starts by drawing mainly authors from Tyrol and increasingly becomes a phenomenon involving nearly all German-speaking regions, whereas the last issues are produced by a small circle of authors. Overall, the *Brenner* is considered to spatially incline towards the cultural centers of that time and especially Vienna (Brenner-Archiv, 2010).

The present object of scientific inquiry does not simply reside in linking text to space, it is tightly linked to the interpretation of texts and maps. Even if the methodology conveys a feeling of scientific objectivity, the validity of mental and computerized operations described here should always be examined with respect to their relevance. Geospatial analysis and spatial representation may indeed be deficient or inadequate. The anthropological significance of toponyms has been emphasized by testimonies gathered on the field (Feld, 1996), but the symbolic role and the expressive power of placenames do not necessarily coincide with Western instrumental science and cartography, in that particular case the world geodetic system and the chosen map projection. In this sense, putting *Der*

---

[1] «Sans qu'il puisse être question d'apporter une réponse à cette énigme que désigne dans notre conscience ou notre inconscience le nom de Rome, de rendre compte même grossièrement de ce foyer capital d'émerveillements et d'obscurités.» Michel Butor, *La Modification*, Minuit, 1957, p. 229.

Brenner on the map implies to put in on the dissecting table and bears both a thrill and a risk (Gerlach, 2014): "adding more to the world through abstraction" and "adding to the riskiness of cartographic politics by proliferating yet more renders of the world".

3.2   Tooling

In order to properly observe this research object in a text collection, a proper scientific apparatus has to be put in place. The specific nature of historical literary corpora calls for ad hoc solutions. Three common issues in geographic information extraction are addressed here: detecting geographical references, disambiguating placenames and developing effective user interfaces. Additionally, the approach described here involves the gathering and curation of supplementary information applicable to historical texts.

The extraction of spatial entities or geoparsing can be considered as one of the most important part in spatial analysis (Fize, Shrivastava, & Ménard, 2017) It has been shown that drawing information from several sources and planning several corresponding phases is useful for gazetteer creation, as a method for automatically constructing a geographic gazetteer using heterogeneous sources of information available on the Internet, can mitigate coverage issues by combining information as well as improve the resolution and the correctness by validating the datasets against another. It is for example useful to mine Wikipedia for location information (Popescu, Grefenstette, & Moëllic, 2008), which results in the constitution of minimum explicit tuples for geographic names: Entity-Name, Entity-Coordinates, and Entity-Type.

The difficulties related to German include capitalization, genitive forms (typically formed with s), adjectival use (frequently as modifiers), and the fact that inhabited placenames are also frequently found as personal names (Volk & Clematide, 2001). These reasons make it difficult to build a gazetteer and prompt for the integration of morphological information. As a general principle, it is expected that "named entity recognition in German will profit from precompiled lists as well as from learning and filtering" (*ibid.*). Indeed, especially for historical texts it is necessary to refine or complement the resources at hand: during the 20th century there have been significant political changes in Central Europe that have severely affected toponyms, so that geographical databases lack coverage and detail.

The toolbox used for the experiments below (Barbaresi, 2017) is currently being developed with historical texts in mind (Barbaresi, 2018b). Its release under an open-source license is meant to ensure reproducibility and extendability in an open science perspective and can hopefully respond to a growing demand in this field.[2] The operations performed prior to the extraction comprise the bootstrapping, filtering and merging of registers. Both the extraction and the disambiguation are parametrized to suit the particular case at hand. As there is no commonly adopted standard for gazetteers they have to be combined. Consequently, the approach employed here allows for additional input, special sorting and prioritized merging, for example to put historical variants in the foreground. Second, it includes helpers to bootstrap geographical data, as knowledge-based methods using fine-grained data improve the results (Vrandečić & Krötzsch, 2014). So far, import filters for GeoNames[3] and structured data from Wikipedia and Wikidata[4] are implemented, with a particular emphasis on data cleaning. Third, an additional layer allows to bypass geocoding for targeted, easily extensible user lists which can operate on token or lemma level (using either linguistic processing or regular expressions and wildcards).

Regarding toponym resolution, two different types of disambiguation methods (Buscaldi, 2011) are included so far in the toolbox: map-based and knowledge-based. It has been shown that an acceptable

---

[2] https://github.com/adbar/geokelone
[3] http://www.geonames.org a project of Unxos GmbH whose main source is the NGA (National Geospatial-Intelligence Agency, USA).
[4] Respectively wikipedia.org and wikidata.org, both being projects of the Wikimedia foundation.

precision can be reached by including meta-information (Pouliquen et al., 2006), which consists here in distance (based on a calculation relative to a contextual setting), type and importance of the entries (as known from information extracted from GeoNames or Wikipedia), as well as immediate context (e.g. the expected range and the last country seen). The process is controlled by parameters such as distance calculations, filter level or size of the search area.

Finally, the results are projected on customized maps, as the toolbox integrates its own visualization component. It is profitable to allow for adaptability of projection and design and to leave it open to the user to refine the map, in a particular emphasis on the concept of visualization.

4 Results

The present experiments use the same text base from the digital edition of the work (AAC-Brenner, Biber, Breiteneder, Kabas, & Mörth, n.d.).

The streamlined process from text to map involves a series of decisions as well as a critical reading of texts and maps. As user-definable settings make results vary, experiments can lead to diverging realizations. The filtering level affects both the construction of gazetteers prior to geoparsing and the toponym recognition phase in itself. Its purpose is to allow for a tighter or looser control on the data, with either restricted options or opportunistic search. The disambiguation phase is controlled by map-based parameters, notably the reference point (set here in Vienna) for distance calculations and the countries in the vicinity, which help identifying the most probable candidates. The cartographic processing in itself does not depict a neutral point of view but an active exploratory process.

Figure 1: Results found throughout the collection with a minimum of filtering

Figure 2: Similar results after restrictive filtering

In order to better assess the impact of filtering and complementary registers, I present and discuss two different comparisons on close reading and on distant reading levels. Specially curated gazetteers are used, while current geographical information is used as a fallback, entries corresponding to European locations are retrieved and preprocessed. Figure 1 displays the results with a minimum filtering on a map showing most of continental Europe. Clusters can be found everywhere, not all of them being either intuitively explainable or justified with respect to the texts. In fact, the map tells more about the gazetteers used for geoparsing as about the work in itself. Current boundaries are retraceable, especially for Germany and Austria, and numerous false positives come from plurilingual countries such as Switzerland or Belgium which are then overrepresented on the map. Figure 2 consists of a similar map featuring the results of maximum filtering level both during the construction of resources and during the extraction process. The map is more easily readable and depicts an accurate centering on Austria and

Western Europe. The overall Westward tropism of the mapped locations seems to coincide with the texts. This map is thus well-suited for further analyses.

Figure 3: Distribution of place names at the
beginning of publication (1910-1912)

Figure 4: Distribution of placenames
in the 1920s (1923-1927)

In a second experiment, filters are used to compare two distinct series of issues. The filtering of result is set to maximum and a time frame is determined. Figure 3 depicts results from the beginning of publication, as issues from 1910 to 1912 have been selected, whereas Figure 4 shows results from the middle of publication in the 1920s when a topical and auctorial shift has taken place. More years are taken into consideration (1923 to 1927) as in the first case, since the journal is more sparsely published and since these years give a good example of the trends during this decade.

Considering that place is "as much a topical consideration as it is a geographical one" (Wrisley, 2017) leads to an interpretation of the maps as topical indexes of the literary work. In this regard, the distributions found in Figures 3 and 4 are similar overall, which highlights a greater spatio-topical continuity than could be expected. In the last case the focus on Tyrol is still present although the number of results is smaller, and Berlin is more prominent as a major cultural center of the time. The noticeable shift from Western to Eastern Europe is a result of the change of focus, the comparatively small number of points to be drawn may corroborate the observations by close readers noting a turn toward more abstract topics in the course of time. The progressive focus on major locations is a phenomenon which is also found in the closely related *Die Fackel*, published by Karl Kraus in Vienna, which echoes the cultural and political reconfiguration of Europe (Barbaresi, 2018c).

5   Discussion

The status of placenames that are to be found and projected on the map ought to be discussed since there are consubstantial ambiguities on linguistic level that complicate the search (Smith & Mann, 2003). In general, successful detection and disambiguation relies on a smart interplay of resources and tools at different levels. Last, the case of either imprecise, vague or vernacular names (Jones, Purves, Clough, & Joho, 2008) is a prominently linguistic issue which can at least be addressed by manual curation and should in any case be attended to. GeoNames, arguably the most commonly used gazetteer, has to be put under scrutiny, as the entries and their classification are subject to numerous problems, mostly unevenly distributed data and sparse metadata, which impact both detection and disambiguation of placenames (Acheson, De Sabbata, & Purves, 2017). Nevertheless, this resource is still valuable mostly because of its coverage of language variants and thus potentially historical variants.

Additionally, as in any kind of linguistic analysis, the research question is only valid if there are significant phenomena to be found in the text. Concerning spatial analysis and placenames the problem

can be more subtle, as the presence of identifiable toponyms is not necessarily conclusive, despite their apparent meaning and the exact linking to a gazetteer. There are "less identifiable" places which are more a "grouping of concepts".[5] Let us assume for now that the placenames found in early 20th century Europe are not void of signification. Nevertheless, the design and selection steps leading to a map are all "inherently rhetorical" (Harley, 1989), they are both the foundation and the product of scientific reasoning and they form a "cultural text" (*ibid.*). There is a necessary distance to be taken in distant or superficial reading, as the product of machine reading can diverge from the expectations. One must bear in mind that "computers are lousy readers" and that "our current digital tools walk a delicate line between analytical power and accessibility" (Wilkens, 2011) Indeed, geoparsing (especially geocoding) implies a necessary reduction to monosemic interpretations, missing for example the vicinity of *Brenner* and *Fackel* as pertaining to the semantic field of fire.

6   Conclusion

This article introduced theoretical and practical instruments combining philological knowledge, geographic information retrieval and visualization, along with the particular example of the detection and projection of spatial patterns in a historical text collection, *Der Brenner*. Two different filter settings as well as maps focusing on the 1910s and 1920s have been compared. In a linguistic perspective, the tools allow for the systematization of research, for a critical approach to the extraction, and for an examination of the very concept of placenames. Examples of issues related to historical text corpora have been discussed, they show the advantages of an appropriate methodology and subsequent data cleaning and filtering.

Drawing maps to study texts leads to yet another level of textuality. By accepting it, "we are able to embrace a number of different interpretative possibilities" (Harley, 1989). Rapidly spotting problems in methodology or datasets, customizing gazetteers and maps can hopefully help mitigating the side effects of large-scale analysis and distant reading. The difference between mere data wrangling and research in digital literary cartography resides precisely in the number and diversity of conceptual and technical filters which are repeatedly applied, consciously or sometimes unknowingly. The chosen approach and its inevitable imperfections have to be brought to light, documented and criticized, as the mapping process represents an interpretation in its own right: "We organize information on maps in order to see our knowledge in a new way." (Turchi, 2007) The results lead us to "ask more questions, consider other possibilities" (*ibid.*) As quantitative and qualitative analysis can go hand in hand, digital literary studies are not mere numeric accounts. They are first and foremost a discovery process, especially considering the complex, intricate nature of such a periodical, which yields maps closer to a deep, ramified rhizome than to a flat constellation (Barbaresi, 2018). In this case, text visualizations are the substrate of interpretable representations which do not follow the underlying data but rather confront them by putting them in perspective.

References

Acheson, E., De Sabbata, S., & Purves, R. S. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, *64*, 309–320.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh geoparser for historical georeferencing. *International Journal of Humanities and Arts Computing*, *9*(1), 15–35.

---

[5] "Like many named places in California it was less an identifiable city than a grouping of concepts -- census tracts, special purpose bond-issue districts, shopping nuclei, all overlaid with access roads to its own freeway." Thomas Pynchon, *The Crying of Lot 49*, 1966.

Barbaresi, A. (2017). Towards a Toolbox to Map Historical Text Collections. In *Proceedings of the 11th Workshop on Geographic Information Retrieval*. ACM. https://doi.org/10.1145/3155902.3155905

Barbaresi, A. (2018). A constellation and a rhizome: two studies on toponyms in literary texts. In B. Noah & K. Marc (Eds.), *Visual Linguistics*. Heidelberg: Heidelberg University Publishing.

Barbaresi, A. (2018b). Borderlands of text mapping: Experiments on Fontane's Brandenburg. In M. Burghardt & C. Müller-Birn (Eds.), *INF-DH-2018*. Bonn: Gesellschaft für Informatik e.V.

Barbaresi, A. (2018c). Toponyms as Entry Points into a Digital Edition: Mapping Die Fackel. *Open Information Science*, 2(1), 23-33.

Biber, H., Breiteneder, E., Kabas, H., & Mörth, K. (n.d.). *Austrian Academy Corpus AAC–FACKEL.*

Borin, L., Dannélls, D., & Olsson, L.-J. (2014). Geographic visualization of place names in Swedish literary texts. *Literary and Linguistic Computing*, 29(3), 400–404.

Brenner-Archiv, F. (Ed.). (2010). *Zeitmesser – 100 Jahre "Brenner."* Innsbruck University Press.

Buscaldi, D. (2011). Approaches to disambiguating toponyms. *Sigspatial Special*, 3(2), 16–19.

Cooper, D. (2017). Digital Literary Cartographies: Mapping British Romanticism. *The Routledge Handbook of Literature and Space*, 135–147.

Feld, S. (1996). Waterfalls of song: An acoustemology of place resounding in Bosavi, Papua New Guinea. In S. Feld & K. H. Basso (Eds.), *Senses of place* (pp. 91–135). School of American Research Press.

Fize, J., Shrivastava, G., & Ménard, P. A. (2017). Geodict: an integrated gazetteer. In *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*.

Gerlach, J. (2014). Lines, contours and legends: Coordinates for vernacular mapping. *Progress in Human Geography*, 38(1), 22–39.

Gumbrecht, H. U. (2009). A Negative Anthropology of Globalization. In *The Multiple Faces of Globalization* (pp. 230–241). Madrid: BBVA, Francisco González et al.

Harley, J. B. (1989). Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 26(2), 1–20.

Hill, L. (2000). Core elements of digital gazetteers: placenames, categories, and footprints. *Research and Advanced Technology for Digital Libraries*, 280–290.

Jones, C. B., Purves, R. S., Clough, P. D., & Joho, H. (2008). Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science*, 22(10), 1045–1065.

Melo, F., & Martins, B. (2017). Automated Geocoding of Textual Documents: A Survey of Current Approaches. *Transactions in GIS*, 21(1), 3–38.

Nouvel, D., Ehrmann, M., & Rosset, S. (2015). *Les entités nommées pour le traitement automatique des langues*. ISTE editions.

Plini, P., Di Franco, S., & Salvatori, R. (2016). One name one place? Dealing with toponyms in WWI. *GeoJournal*, 1–13.

Popescu, A., Grefenstette, G., & Moëllic, P. A. (2008). Gazetiki: automatic creation of a geographical gazetteer. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 85–93). ACM.

Pouliquen, B. (2006). Geocoding multilingual texts: Recognition, disambiguation and visualisation. In *Proceedings of LREC* (pp. 53–58). ELRA.

Ramsay, S. (2005). In Praise of Pattern. *TEXT Technology: The Journal of Computer Text Processing*, 92(2), 177–190.

Smith, D. A., & Mann, G. S. (2003). Bootstrapping toponym classifiers. In *Proceedings of the HLT-NAACL 2003 workshop on Analysis of Geographic References* (pp. 45–49). Association for Computational Linguistics.

Southall, H., Mostern, R., & Berman, M. L. (2011). On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5(2), 127–145.

Turchi, P. (2007). *Maps of the Imagination: The Writer as Cartographer*. Trinity University Press.

Volk, M., & Clematide, S. (2001). Learn-Filter-Apply-Forget. Mixed Approaches to Named Entity Recognition. In *Proceedings of NLDB* (pp. 153–163).

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a Free Collaborative Knowledge Base. *Communications of the ACM*, 57(10), 78–85.

Westphal, B. (2014). Foreword. In R. T. J. Tally (Ed.), *Geocritical Explorations : Space, Place, and Mapping in Literary and Cultural Studies* (pp. ix–xv). Palgrave Macmillan.

Wilkens, M. (2011). *Contemporary Fiction by the Numbers*.

Wrisley, D. J. (2017). Locating Medieval French, or Why We Collect and Visualize the Geographic Information of Texts. *Speculum*, *92*(1), 145–S169.