

Adrien Barbaresi

A Constellation and a Rhizome: Two Studies on Toponyms in Literary Texts

Abstract Although the attention of linguists is commonly drawn to forms other than proper nouns, the significance of place names in particular exceeds the usual frame of deictic and indexical functions, as they encapsulate more than a mere reference in space. In this article, two different examples are presented in order to understand them, along with two practical examples of visualization in literary texts from the beginning of the 20th century. Research on toponym extraction and linkage is discussed from an interdisciplinary perspective, as digital literary studies are not mere numeric accounts: if they deal with detecting, counting, and projecting occurrences, one also ought to describe and criticize the detachment provoked by “blind” computer-based thinking.

The first case consists of a preliminary study of travel literature based on Richthofen’s *Travel Journals from China* (1907). The resulting map retraces the path taken by the author in the Shandong province by combining coordinates, sequences, and a sense of time. In order to critically analyze this synthesis the concept of constellation is introduced and discussed. The second study focuses on a complex work, the literary magazine *Die Fackel* (1899–1936) with the particular example of co-occurrences of toponyms. The paths drawn on the map depict chains of thought and lines of flight. The result is understood through the concept of the rhizome, by which heterogeneous information can be connected and displayed. The finality of Visual Linguistics does not reside in an apparatus but rather in the substrate of interpretable representations which put words into perspective.

1. Introduction

Although the attention of linguists is commonly drawn to forms other than proper nouns, the significance of place names in particular exceeds the usual frame of deictic and indexical functions, as they embrace more than a mere reference in space. The value of toponyms is highlighted in testimonies gathered in the field, such as the comment by Ulahi from Papua New Guinea who asks: “I

haven't heard your land names, so who are you?" His study on Bosavi language leads Feld (1996) to the idea that "place significance neither starts nor ends with the linguistic referentiality of placenames", and that "there is considerable variation in how names hold and unleash significance".

In Western tradition, a current of reflection, whose origin can be dated back to the 1960s, has provided the theoretical foundations of the *spatial turn*, whose epitome is the concept of space as emergent rather than existing a priori, and composed of relations rather than structures (Warf 2009). As Foucault described in a seminal lecture:

"Nous sommes à un moment où le monde s'éprouve, je crois, moins comme une grande vie qui se développerait à travers le temps que comme un réseau qui relie des points et qui entrecroise son écheveau."¹ (Foucault [1967] 1984)

Foucault further explained that "our epoch is one in which space takes for us the form of relations among sites". As a consequence, both the definition and importance of space have been re-evaluated throughout the humanities.

More recently, the emergence of *GeoHumanities* (Dear et al. 2011) or *Spatial Humanities* (Bodenhammer et al. 2010), has prompted the exchange of research objects between disciplines as well as enforcement of the spatial turn in practice through specific methods of analysis. Although some observe that there is "remarkably little overlap between the digital humanities community and the spatial history community" (Mostern & Gainor 2013), the common denominator seems to be the will to open up new spaces and experiment in a transdisciplinary perspective:

"The parallel disciplinary structure of German-language *Literaturwissenschaft* and linguistics is not merely anecdotal; rather, it is an index of the necessary interchange of methodologies and content which holds beneficial possibilities for both fields." (Domínguez 2011)

Following these premises, I wish to present two studies that center on the visualization of place names in literary texts, with particular emphasis on the concept

1 "We are at a moment, I believe, when our experience of the world is less that of a long life developing through time than that of a network that connects points and intersects with its own skein." <http://foucault.info/doc/documents/heterotopia/foucault-heterotopia-en.html>

of visualization, that is on the processes and not on the products (Crampton 2001). As a consequence, I feel the need to present them with a much needed critical apparatus, by giving a theoretical perspective on what is being shown and seen: Firstly because digital methods in humanities ought to be criticized (Wulfman 2014) and secondly, because the cartographic enterprise bears both a thrill and a risk: on the one hand “adding more to the world through abstraction”, and on the other hand “adding to the riskiness of cartographic politics by proliferating yet more renders of the world” (Gerlach 2014).

The remainder of this article details theory and practice with the following structure: First, research on toponym extraction and linkage is discussed from an interdisciplinary perspective; then the constellation paradigm is presented along with a practical study on travel literature; finally, the notion of the rhizome is introduced and presented by means of a diachronic study.

2. Drawing points and lines out of words

2.1 Distant cartographic reading on relative maps

Progress in fulltext geocoding, also known as “geoparsing” or “geographic information retrieval” (Leetaru 2012), is tightly linked to progress in mapping systems, mostly thanks to a technology-driven evolution (Juvan 2015) as Geographic Information Systems (GIS) and series of tools come from other disciplines. An underlying assumption resides in the belief that understanding language and literature is not accomplished by studying individual texts, but by aggregating and analyzing massive amounts of data (Jockers 2013). Because it is impossible for individuals to “read” everything in a large corpus, advocates of distant reading employ computational techniques to “mine” the texts for significant patterns and then use statistical analysis to make statements about those patterns (Wulfman 2014). One such is Moretti (1999) who pleads for “a geography of literature” and “distant cartographic reading”.

There is however a notable difference between literary geography and literary cartography, the first one being essentially anecdotal and auxiliary, if not ancillary (Juvan 2015). Although literary geography has been existing for more than 100 years (Piatti et al., 2011), much work still has to be done to define and establish literary cartography, an interdisciplinary field which has been evolving at an exponential pace over the last decade (Caquard & Cartwright 2014).

Concerning the maps themselves, the consensus in the research community has evolved towards a relativity in construction and uses of maps: “post-representational cartography” (Rossetto 2014), where there is neither a “ground truth”, nor a “cartographic truth”, and where “the map is not objectively ‘above’

or ‘beyond’ that which is represented” (Crampton 2001). Although the maps seem immediately interpretable, they are not an objective result but a construct, the result of filtering, “a connection made visible” (Moretti 1999). As such, cartography is not the realization of static maps, but rather the description of emergent structures, and there is no single or best map. The paradigm of *geographic visualization* stands in opposition to the tradition:

“Traditional cartography has emphasized public use, low interactivity and revealing knowns, while visualization emphasizes private use, high interactivity and exploring unknowns.”
(Crampton 2001)

In literary studies, it is understood that maps are only the beginning of exploratory work, not only following from said paradigm but also out of a defiance against quantitative information of an additive, cumulative nature about potentially heterogeneous phenomena:

“Whenever literary scholars screen, read, interpret and compare the maps, they do what is regarded as one of their core competences: to consider carefully ambiguities, to compare, to contextualise, to shed light on historical references, to juxtapose several readings, to combine methods and tools.” (Piatti et al. 2011)

Counting words, in this particular case place names, does not appear to be enough for the researcher in literary studies. Even if the map in itself is relative, being “less important than the process of making it and using it” (Caquard & Cartwright 2014), it plays an ambiguous part in distant reading, since it has to be flexible enough to adapt to new contexts and analyses, while remaining exact and in this sense trustworthy. The information it contains and reveals cannot be verified on a point-per-point basis, yet it can be the starting point of a comprehensive interpretation. For this reason, the ability to detect and project place names with reasonable accuracy is paramount.

2.2 Placing points: On the extraction of toponyms

In the field of information extraction or information retrieval, named entity recognition is a set of text mining techniques designed to discover named entities, connections and the types of relations between them (Chinchor 1997). The particular task of finding place names in texts is commonly named place names extraction or toponym resolution. It involves the detection of words and phrases that may potentially be proper nouns as well as a second operation classifying them as geographic references (Nouvel et al. 2015). A further step, geocoding, resides in disambiguating and adding geographical coordinates to a place name:

“At its core, fulltext geocoding involves scanning a body of text to identify potential geographic references and then using an external knowledgebase, called a ‘gazetteer’, and document context to disambiguate and convert the references to a geo-spatial form.” (Leetaru 2012)

Named entity resolution often relies on named-entity recognition and artificial intelligence (Leidner & Lieberman 2011). However, knowledge-based methods using fine-grained data, for example from Wikipedia, have already been used with encouraging results (Hu et al. 2014). Work by Efremova et al. (2015) on family relationship extraction from historical documents proves that it is possible to use a name dictionary as well as patterns to perform the extraction and remove ambiguity in a robust way, although the documents span around 500 years. Pouliquen et al. (2006) demonstrate that an acceptable precision in the detection and disambiguation of place names (76 % in their study) can be reached by including information such as distance, importance, immediate lexical context, and main places of the text to be analyzed.

The first study below relies on manually annotated data, while the second is based on automatic extraction. One of their common denominators is the necessity to construct appropriate gazetteers, that is mapping historical and possibly not yet standardized variants to a canonical standard from which to derive geographic metadata. Another common denominator consists of the connections that are drawn between the extracted toponyms in the exploratory sense of geographic visualization.

2.3 Connecting and dividing: the ambiguous nature of lines

Maps are mostly projected in Euclidean spaces where two points are connected by a single line. By extension, the word *line* defines a series of connected points on a plane, potentially both linking and separating. Lines can enforce and divide when maps are used as instruments of power, which is another reason why the proponents of post-representational cartography call for a change of perspective. For example, lines that draw state boundaries are increasingly called into question, especially concerning historical states and texts, as they fix an evolving process and convey a sense of immobility that fails to describe the past accurately:

“We need to recognize that territorial maps of ancient states are an idealized projection of state authority rather than a depiction of the way in which ancient political domains were actually governed.” (Smith 2005)

In the framework of geographic visualization, advances have been made towards less static maps, with the wish to foster more flexibility on the map and in the mind of the reader, allowing for reflection out of the box and outside of the boundaries:

“Recent attention to globalization, diaspora, ‘nomadism’, and cyberspace is showing us the need for new and powerful theoretical work to replace, rather than simply supplement, the polemics and models produced by an academic collectivity concerned mostly with locatable cultures, bounded nations, and the imperial past” (Campbell 2002).

What is true for the territory also holds true for the texts, linkages can be seen as “mappings and tracing imposed on the data” (Wulfman 2014). The lines may also reveal spatial patterns that would otherwise remain hidden in texts (Bodenhauer et al. 2010), or connect ideas, places, or peoples. As well as lines of force, there can also be lines of thought, traces figuring steps of reflection and analysis. This ambiguous nature calls for a differentiated approach. As long as the traces can be flattened, recombined, and superposed (Latour 1985), they allow for a scientific process of (re-)construction which is open to additions, changes of scale and perspective.

Visualization of linguistic phenomena has to account for their changing nature relative to context and passage of time, thus going deeper than a mere

graph, a general and operative concept which usually implies a series of operations performed upon it. Beyond the immediacy of nodes and edges, the relations may be expressed through the concept of network. However, precisely in the case of connections, the word “network” is to be used with caution as Latour (1999) suggests. Although it is ubiquitous in the terminology of the spatial turn, the now predominant interpretation in the sense of the World Wide Web, suggests an immediacy which is contrary to the status it had before. In most occurrences of the term, “meshwork” would be more appropriated than “network” (Ingold 2007). In the following, two different modes of analysis featuring points and lines are presented. The potential superficiality of networks is carefully avoided as two related alternative concepts which both break the linear model of reflection (Wu 2009) are presented: the constellation and the rhizome.

3. First study: Constellation

3.1 The concept of constellation

According to the words of De Certeau (1990), where the map splits, the narration – as diegesis – traverses. Even if a map remains static in comparison to the movement of the diegesis, an attempt can be made to try and overcome this dichotomy. In this first study, a decision has to be made about whether to trace the line or not, which makes it a prototypical constellation, an assemblage of points which let figures and interpretations emerge:

“As we make constellations by picking out and putting together certain stars rather than others, so we make stars by drawing certain boundaries rather than others. Nothing dictates whether the skies shall be marked off into constellations or other objects.” (Goodman 1983)

Stars and skies are here place names and come from the spatio-temporal frame of the narration in travel literature, a specific case which challenges the idea that “linkages are not in the data” (Wulfman 2014). In fact, it is possible to try to visualize a progression in the narration by bringing phenomena to light with annotation, and drawing lines between certain points to make patterns visible which would remain unnoticed otherwise.

3.2 Modus operandi

The maps with an “indexical function” (Juvan 2015), are part of a preparatory study on travel literature, which includes the annotation of place names from Richthofen’s *Travel Journals from China* (1907) in XML format and conform to the guidelines of the Text Encoding Initiative.² The outline below focuses on four weeks of travel in 1869 through the Shandong province (East China). Annotation of toponyms is done manually due to difficulties with non-standard transcriptions of historical Chinese names.³ For each identified name, Wikidata is used as reference. This is a document-oriented, collaboratively edited knowledge base operated by the Wikimedia Foundation (Vrandečić & Krötzsch 2014). As it is international and editable, it allows for the registration of linguistic variants, and metadata such as coordinates can be added. The *type* attribute defines whether the given location is to be represented as a point or as a surface. XML TEI attributes are used to encode direction of travel and sequences of visited places (*next* and *prev*), as well as information about time series (*when* and *notafter*). Finally, toponyms and metadata are extracted from the resulting XML document and combined with information from the authority file before being projected on a map.

3.3 Result

The result combines coordinates, sequences, and a sense of time, which are depicted on Figure 1 by (respectively) points, lines, and a color scheme, projected on a map using the cartographic software TileMill⁴ and customized with CartoCSS. For the sake of clarity, current standard names in English are also projected on the map, as well as historical borders of Chinese provinces in the 19th century.⁵ Additionally, a color contrast distinguishes the land mass (Mainland China) from the sea (Yellow Sea and East China Sea). The size of the dots is in relation to toponym frequency in the text. Different shades of red illustrate time differences from the first to the fourth week of the trip, while gray points depict places which are named without having been visited., and another constellation, whose lines are not retraced, contrary to the itinerary.

² <http://tei-c.org>

³ The preliminary steps have been performed jointly with Benno Wagner (Zhejiang University) and Li Liu (Stuttgart University).

⁴ <https://github.com/mapbox/tilemill>

⁵ The source is a prototypical spatial humanities project, China Historical GIS: <http://www.fas.harvard.edu/~chgis/>