The Mathematical Structure of Geographic Description

MICHAEL GAVIN

"The seventeenth-century English textual landscape," writes Cynthia Wall, "was well populated with the praxes of textual description."[1] Indeed, geographical accounts of the British Isles flourished during the period. William Camden's magisterial chorography, *Britannia* (1586), first translated into English in 1610, gathered detailed maps and prose descriptions of English towns, cities, and parishes. His work was reprinted, adapted, and imitated throughout the century. John Adams's extraordinary *Index Villaris* (1690) estimated the locations and basic demography of 24,000 English towns, using the latest techniques of analytical geometry for projecting the curvature of the Earth onto a two-dimensional cartographic plane. Geographical descriptions were not limited to England, however. They roamed over all parts of the known world: topographical dictionaries contained fantastical accounts of Asia, America, and Africa, as well as of the Arctic and Antarctic zones. Textbooks like *Cosmographia* (1679) and *Geography Anatomiz'd* (1699) provided chorographies of the world while instructing readers to use globes and maps to find places and to calculate relations among them. More practical reference manuals like James Wadsworth's *Evropean Mercury* (1641), John Ogilby's *Book of Roads* (1675), Thomas de Laune's *Present State of London* (1681), and Laurence Echard's *Newsman's Interpreter* (1692) mapped places along spatial networks by describing roads, postal services, and stage-coach routes that connected London to cities across the British Isles, Europe, and Asia.

The "English textual landscape," to return to Wall's phrase, was not a landscape at all, but an ever-growing body of descriptive prose oriented globally and organized mathematically. Geographical description differs from other language by relying explicitly on an underlying spatial model. Like other words, toponyms acquire meaning through differentiation. "Worcestershire" was used differently from "Aberdeenshire," and "England" was used differently from "Madagascar." But they also depend for their meaning on a sense of place that separates them spatially: provinces of France are close; the Arctic and Antarctic zones are far apart. How were such distances conceptualized? The geodesic distance that separates two places could be estimated by applying trigonometry to a model of the Earth's shape, and geographical dictionaries often included coordinates of latitude and longitude along with instructions for performing such calculations. Adams's *Index Villaris* accompanied a large atlas of England that located its places within a network of spatial distances. The travel distance that separates two places could also be estimated by measuring paths over a network of roads or rivers. Ingolstadt, Vienna, and Budapest could all be reached by floating the Danube, while a journey from Paris to Rome involved many stops and required changing carriages several times. And at its most basic, chorography grouped toponyms into categories resembling tree-shaped directed graphs in which the globe was divided into continents, continents into nations, nations into regions, and so on.

This paper combines corpus linguistics, geographic information science (GIS),

---

[1] Cynthia Wall, *The Prose of Things: The Transformation of Description in the Eighteenth Century* (University of Chicago Press, 2006), 41.

and network analysis to describe the mathematical foundations of seventeenth-century geographical discourse. Working from a small corpus of chorographies, textbooks, and travel guides taken from the *Early English Books Online* collection, I have created three tightly related datasets.

The first is a collection of several thousand place descriptions, organized by toponym. The second is a geospatial database that contains the latitude and longitude of each place and that uses custom functions to measure geodesic distance according to methods prescribed in seventeenth-century reference books. The third is a network model of toponyms, showing how each document in the corpus organizes places in relation to each other. Across these models, measure the *distribution of geographic diction*; that is, I compare and contrast the vocabulary of place descriptions in the corpus, and I evaluate similarities and differences in the language with proximities and distances over the geospatial models.

My immediate aim is to test the following hypothesis: If geographical description is meaningfully structured by underlying spatial models, similarity in diction should tend to correspond to proximity in space.[2] My analysis shows that this hypothesis largely holds, but that "proximity in space" depended on historically contingent frames of reference. In seventeenth-century discourse, China and Africa, for example, were further from England than France was, both spatially and semantically, but they were described using a common language of exoticism and so were more similar to each other than would be predicted based on the geodesic distance that separated them. In general, I expect that network-path distance will be better than geodesic distance at predicting semantic difference: places that were connected in physical or conceptual networks had more in common, in the seventeenth century, than places that were merely proximate on the globe. However, language remained strongly geo-correlative, and the *Early English Books On-Line* collection distributes meanings into geographic regions with a clarity of distinction that is, perhaps, surprising.

My larger goal is to invite scholars of eighteenth-century literature and culture to take another look at space theory and, in particular, to introduce them to theoretical traditions in mathematics, information science, and geography that are newly relevant to our field. Statistical models of word meaning were first designed in the 1950s (though linguists at the time lacked the requisite data or computing power), and geography was among the first disciplines, in the 1960s, to reconceive its subject by incorporating the mathematical innovations of graph theory.[3] Even in the seventeenth century, geogra-

---

[2] Geospatial semantics thus combines the "distributional hypothesis" of computational semantics with the principle of *spatial autocorrelation* in geography. See Zellig Harris, "Distributional Structure," *Word* 10 (1954), 146-162 and Luc Anselin, "Spatial Econometrics," in Badi Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell, 2001).

[3] For computational semantics, see Harris (1954), Warren Weaver, "Machine Translation" (1955) in S. Nirenburg et.al., eds., *Readings in Machine Translation* (MIT Press, 2003), J. R. Firth, *Papers in Linguistics* (Oxford, 1957). For the mathematical foundation of graph theory, see Oystein Ore, *Theory of Graphs* (American Mathematical Society, 1962) and Frank Harary, et. al., *Structural Models* (Wiley, 1965), and for its application to geography see Peter Haggett and Richard Chorley, *Network Analysis in Geography* (Edward Arnold, 1969), Marc Barthélemy, "Spatial Networks," *Physics Reports* 499 (2011), 1-101, and Rob Shields, "Cultural Topology: The Seven Bridges of Königsburg," *Theory, Culture, & Society* 29:4/5 (2012), 43-57. The basic principles of

phy was the first of the human sciences to rely explicitly on mathematics: circumnavigation of the Earth popularized three-dimensional spatial models that required fairly complex geometric solutions. Much of my presentation will be devoted to narrating these analogous histories. Now, with the advent of full-text corpora like *EEBO* and *ECCO*, a similar transformation is underway in literary and cultural studies.[4] This change is often understood as a change in scale, but my presentation suggests a different viewpoint.[5] Quantitative methods don't require big data. Instead, they require creative theorization.

Geographic information science treats space as a complexly layered, multi-dimensional object in which the distance between any two points is not fixed but subject to varying description and continuous change. Geographers of the seventeenth century knew, too, that space was nothing so easily thinkable as a landscape. Tracing the spatial distribution of historical textualities requires of cultural history a similarly rigorous theory.

GIS and quantitative geography are explained in Fotheringham, Brunsdon, and Charlton, *Quantitative Geography Perspectives on Spatial Data Analysis* (SAGE Publications, 2000) and Ian Gregory and Paul Ell, *Historical GIS* (Cambridge, 2007).

[4] *Eighteenth-Century Collections Online* is still largely closed to scholars, but the public release of *Early English Books Online* documents by the Text Creation Partnership in 2015 has stimulated significant advances in computational scholarship. See, for example, Laura Estill, Diane Jakacki, and Michael Ullyot, eds., *Early Modern Studies after the Digital Turn* (Iter & ACMRS, 2016). The eighteenth century will soon become a dark age if our archive is not modernized.

[5] For "scale" as the central problematic, see English and Underwood (2016).