Place Reference in Text as a Radial Category:
A Challenge to Spatial Search, Retrieval, and Geographical Information Extraction from Documents that Contain References to Places

ALAN M. MACEACHREN
Professor,Geography
Affiliate Professor, Information Sciences & Technology
Director, GeoVISTA Center
The Pennsylvania State University
Email: maceachren@psu.edu

The concept of place is fundamental to spatial search as it relates to finding geographical information in heterogeneous information repositories as well as to the kinds of "spatial search" that are central to human wayfinding in the world. This position statement focuses on place in the context of spatial search for and within text documents, but the conceptualization of place proposed has implications for spatial search more broadly. Spatial search in text documents requires solving a range of challenges related to place, including: recognizing place entity mentions in text, disambiguating and locating the place entities, and determining whether the document is "about" the place(s) mentioned or those place mentions serve some other purpose.

Research over the past several years has provided a base from which to address each of these questions (Zhang et al. 2009; Zhang et al. 2010; Jaiswal et al. 2011; Zhang et al. 2012; Karimzadeh et al. 2013; Xu et al. 2014; Wallgrün et al. submitted). Based upon that research, I propose that one of the most fundamental issues in spatial search focused on retrieving and extracting geographically-relevant information from documents is understanding and being able to recognize what constitutes a reference to a place. My argument here is that place reference is best conceptualized as a *radial category*, but that current entity recognition approaches take a classical category theory approach to recognizing place references. This existing approach impedes strategies for spatial search in text as well as those for geoparsing and mapping the content of documents.

It is important to distinguish between "place" and "place reference." Conceptualization of place is also a fundamental research question (Goodchild 2011) and category theory (within which the concept of radial categories is defined) has been applied successfully to investigate conceptualization of place (Mark, Smith, and Tversky 1999; Lloyd, Patton, and Cammack 1996). Here, however, my focus is on linguistic or textual references to place and how to recognize and interpret them. Thus, the focus is not on what a place is conceptually or in practice (although conceptualization of place is important in understanding place references) but on distinguishing references to place from references to people, organizations, and other entities.

Radial categories (Lakoff 1987), are structured with respect to a prototype and include members in relation to some "center" or prototype (Janda 2010). While I am not aware of any empirical research focused on identifying prototype place references (there is work on basic level geographic features that includes place concepts, [e.g., Lloyd, Patton, and Cammack 1996]),

nor the characteristics of "place reference" as a category, I can speculate that London or Paris might constitute an exemplar, thus close to a prototype place reference for many people. But, someone from China might center their conceptualization of place on Beijing while someone from India might have New Delhi as their prototype. On the other hand, an Amish farmer from the eastern U.S. might center their concept of place on the home farm or local livestock market. Thus, the category of place differs for individuals from different places and who have a different pattern of behavior in the world. Whatever the prototype, radial categories include entities that are more or less representative, thus more or less central or peripheral to the prototype. The position in this radial category space has implications for strategies used to recognize place references in text and for how such references are tagged for subsequent use.

For the category of "place references" (in text or speech), centrality (and thus the likelihood that a speaker/writer, in a statement containing a place name or related reference, is specifically thinking of an entity as a "place" versus another entity type) is likely to depend upon a range of factors. As part of recent research to develop tools for place reference corpus building and then to apply the tools to building a Twitter place reference corpus (Wallgrün et al. submitted), several factors were identified as responsible for disagreements among human annotators (given the task of tagging place references in tweets). Drawing on this work and further analysis of our tweet-tagging results, I propose the following as factors that influence whether a named entity was intended as a place reference; the combination of factors will determine relative position of a reference within or outside the radial place reference category:

Geographical scale: Research has demonstrated that bonds people form with places differ according to geographical scale (e.g., neighborhood, city, country) (Bernardo and Palma-Oliveira 2013). The intensity of place identity and attachment is likely to influence the extent to which a named entity is conceptualized as a "place."

Use as a noun versus adjective: Names as nouns are more certainly place references than names used as an adjective. But, use as a possessive adjective is intermediate. Given three endings of the phrase "I'm going to {the Pirates game in Pittsburgh; Pittsburgh's Pirates game; the Pittsburgh Pirates game}", the first is clearly a place reference and the last is clearly a reference to a sports team (that is from a place). The middle case, however, can be interpreted as a reference to the place, which has a sports team for which only the nickname is given.

Precision of reference: The distinction between an organization and an instance of that organization that is a place is a function of precision of reference. The statement "I often drink Starbucks coffee" refers to the company while "We usually stop at Starbucks" might be a reference to a particular coffee shop or to the chain generally. In contrast, "let's meet at the Starbucks in the hotel lobby" is more certainly a reference to a (within building) place.

Interdependence: Entities can be clear place references in one context and a qualifier of a separate place reference in another. For example, in the two version of this sentence: "I'm traveling to {Illinois; Springfield, Illinois}, the first refers to the state as a distinct place while the second uses "Illinois" to clarify which Springfield is intended (with Illinois secondary).

Agency or lack of it: Place names can be used in ways that imply agency; in these cases, there is a tension between conceptualizing the reference as a place versus as an agent exhibiting behavior. An example is: "Canada Takes a Wait-And-See Approach to New Cage Regulations," Alberta Farmer, Posted Nov. 7, 2011 by Sheri Monk in Livestock; accessed Sept. 19, 2014

(http://www.albertafarmexpress.ca/2011/11/07/canada-takes-a-waitandsee-approach-to-new-cage-regulations/).

   This statement only scratches the surface of the potential application of cognitive category theory to conceptualization of place in the context of spatial search. There is a range of past research on application of category theory to understanding how geographical features are conceptualized (e.g., Mark, Smith, and Tversky 1999; Usery 1993) and newer work on application to contexts such as tagging Flickr photos (e.g., Stvilia and Jörgensen 2010). A next step is to focus on integrating knowledge about how place and related geographic features are conceptualized with that on how place is referred to in text. This integration has the potential to enhance our ability to search for documents that are "about" places of interest and to recognize, extract, disambiguate, and locate the references. One objective is to develop entity recognition methods that take context into account and that can recognize different kinds and intensities of place reference so that spatial search for geographic information in text documents can be better tailored to particular use contexts.

References:

Bernardo, F., and J. Palma-Oliveira. 2013. Place identity, place attachment and the scale of place: The impact of place salience. *Psyecology* 4(2): 167–193.

Goodchild, M. F. 2011. Formalizing Place in Geographic Information Systems Communities, Neighborhoods, and Health. In *Communities, Neighborhoods, and Health*, eds. L. M. M. Burton, S. A. P. Matthews, M. Leung, S. P. A. Kemp and D. T. T. Takeuchi, 21–33: Springer New York.

Jaiswal, A., X. Zhang, P. Mitra, S. Pezanowski, I. Turton, S. Xu, A. Klippel, and A. M. MacEachren. 2011. GeoCAM: A Geovisual Analytics Workspace to Contextualize and Interpret Statements about Movement. *Journal of Spatial Information Science* 3: 65–101.

Janda, L. A. 2010. Cognitive linguistics in the year 2010. *International Journal of Cognitive Linguistics* 1 (1): 1–30.

Karimzadeh, M., W. Huang, S. Banerjee, J. O. Wallgrün, F. Hardisty, S. Pezanowski, P. Mitra, and A. M. MacEachren. 2013. GeoTxt: A Web API to Leverage Place References in Text In *7th ACM SIGSPATIAL Workshop on Geographic Information Retrieval*. Orlando, FL: ACM.

Lakoff, G. 1987. *Woman, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lloyd, R., D. Patton, and R. Cammack. 1996. Basic-level geographic categories. *Professional Geographer* 48(2): 181–194.

Mark, D. M., B. Smith, and B. Tversky. 1999. Ontology and Geographic Objects: an empirical study of cognitive categorization. Paper read at COSIT'99—Conference on Spatial Information Theory, at Berlin.

Stvilia, B., and C. Jörgensen. 2010. Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology* 61(12): 2477–2489.

Usery, L. 1993. Category theory and the structure of features in geographic information systems. *Cartography and Geographic Information Systems* 20(1): 5–12.

Wallgrün, J. O., M. Karimzadeh, A. M. MacEachren, F. Hardisty, S. Pezanowski, and Y. Ju. submitted. Construction and First Analysis of a Corpus for the Evaluation and Training of Microblog/Twitter Geoparsers.

Xu, S., A. Klippel, A. M. MacEachren, and P. Mitra. 2014. Exploring regional variation in spatial language using spatially-stratified web-sampled route direction documents. *Spatial Cognition and Computation*.

Zhang, X., P. Mitra, A. Jaiswal, S. Xu, A. M. MacEachren, and A. Klippel. 2009. Extracting route directions from webpages. Paper read at Twelfth International Workshop on the Web and Databases (WebDB 2009) June 28, 2009, at Providence, Rhode, Island, USA.

Zhang, X., P. Mitra, A. Klippel, and A. MacEachren. 2010. Automatic Extraction of Destinations, Origins and Route Parts in Human Generated Driving Directions. Paper read at GIScience 2010, at Zurich, Switzerland.

Zhang, X., B. Qiu, S. Xu, P. Mitra, A. Klippel, and A. M. MacEachren. 2012. Disambiguating Road Names in Text Route Descriptions using Exact-All-Hop Shortest Path Algorithm. Paper read at European Conference on Artificial Intelligence (ECAI), Aug. 27–31, at Montpellier, France.