

How Gazetteers Work:

Cultural and Linguistic Influences on the Georeferencing Process

David M. Mark
NCGIA & Geography
University at Buffalo, Buffalo NY, USA

As noted in the workshop introduction, georeferencing by naming is universal, and has existed for a long time. Knowing the names has long been a traditional way of demonstrating ownership of country. Before written language and graphic maps, geographic information was often preserved and transmitted through stories, which often also included other important cultural information or moral codes. Place names, and the places themselves, formed 'retrieval keys' for the associated information. Keith Basso's book, "Wisdom Sits in Places" explores such a system still in use by the Western Apache, and there are many examples from Homer's Iliad and Odyssey and the Viking Sagas to the Navajo origin stories and traditions in many and cultures.

Today, we continue to use placenames in our everyday lives, to communicate about locations. But names also play a critical role in accessing information via the Web. Gazetteers provide a link between place names and locations externally, first in printed forms and more recently using digital computers.

In this paper, I will present some thoughts on the cognitive process of how place names *work*, and how they connect names to locations. I will then outline various places in this process where cultural, linguistic, and national differences influence these processes and provide challenges for multilingual gazetteers.

A gazetteer consists of associations among placenames, categories, and footprints (Hill et al., 1999; Hill, 2000). But each of these components is itself complex. The semiotic triangle (Ogden and Richards 1923) is used here as a key model to provide insight into how meaning and reference work in language in general and in gazetteers in particular. Normally, the semiotic triangle shows how a sign has meaning. The sign or symbol, perhaps a word or a name, is linked by a concept (mental structure) to an instance or category in the world. The link between sign and referent is dashed to illustrate that normally, words are linked to things in the world only through concepts.

For the placename corner of the gazetteer triangle, the operation of the triangle is simple. The proper name, "The Matterhorn" or "Santa Barbara", is the sign. The referent is a particular feature in the environment or landscape. Our concept of the appearance and characteristics of The Matterhorn or Santa Barbara allows a two-way associative link between those.

The category corner of the gazetteer triangle has a similar structure. The sign is a landscape generic or feature code such as 'mountain' or 'city'. The concept is the general idea of a mountain or city in general. And the referent is a set of all mountains or cities or a representative sample.

The footprint corner of the gazetteer triangle is the only one that does not seem to expand into a semiotic triangle. Rather, it needs an idea of how entities of that type are bounded (crisp, graded, what gradient; cf. Montello *et al.*, 2005), and the application of the appropriate boundary concept

to the particular entity being referenced.

Cultural and Linguistic Differences in Placenames, Categories, and (Perhaps) Footprints

It is clear that many of the above components vary by language and by culture. Most of us are familiar with different language-specific placenames. Firenze becomes Florence, London becomes Londres, and Cologne becomes Köln or Colonia; even though these are usually just symbol substitutions and refer to "the same thing", that will not always be the case. That is because categories also vary by culture and language, sometimes by splitting or merging (for example, English 'river' splits (more or less) into 'fleuve' and 'rivière' in French) but often in more complicated ways (Mark, 1993; Mark and Turk 2003; Mark *et al.*, in press). Although geographic entity delimitation has not been studied cross-linguistically, it seems quite possible that the footprint of "the same" feature might be significantly different in different cultural and linguistic contexts, especially for features with graded boundaries such as mountains (Smith and Mark 2003).

The presentation will some examples and suggest some approaches to solving the multilingual and multicultural aspects of gazetteers.

References

- Hill, L., 2000. Core elements of digital gazetteers: placenames, categories, and footprints. Proceedings of the 4th European Conference on Research and Digital Libraries. Lecture Notes in Computer Science 1923, pp. 280.
- Hill, L., Frew, J., and Zheng, Q., 1999. Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library. Digital Library 5(1), Corporation for National Research Initiatives, Technical Report: january99-hill.
- Mark, D. M., 1993. Toward a Theoretical Framework for Geographic Entity Types. In Frank, A. U., and Campari, I, editors, *Spatial Information Theory: A Theoretical Basis for GIS*, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 716, pp. 270-283.
- Mark, D. M., and Turk, A. G., 2003. Landscape Categories in Yindjibarndi: Ontology, Environment, and Language. In Kuhn, W., Worboys, M., and Timpf, S., Editors, *Spatial Information Theory: Foundations of Geographic Information Science*, Berlin: Springer-Verlag, Lecture Notes in Computer Science No. 2825, pp. 31-49.
- Montello, D. R., Goodchild, M. F., Gottsegen, J., and Fohl, P., 2003. Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries. *Spatial Cognition & Computation*, 3(2&3), 185–204.
- Ogden, C. K., and Richards, I. A. 1923. *The Meaning of Meaning*, Harcourt, Brace, and World, New York.
- Smith, B., and Mark, D. M., 2003. Do mountains exist? Towards an ontology of landforms. *Environment and Planning B: Planning and Design*, 30(3), 411-427.