**Related work**

We are interested in the extraction of geographic information in the form of triplets, from unrestricted NL place descriptions. Research in the past had focused on mining specific and application-dependent geographic information from controlled language expressions [9, 16, 22] or from certain contexts, such as route descriptions. Hornsby and Li [\*\*] provided a conceptual framework that could serve as the basis for the tagging and parsing of text to extract movement information, while the semantics of locative expressions of various languages are studied in Kracht *et al*. [\*\*]. Linguistic constructs are often complex and lead to under or over-specifications when trying to match NL to formal frameworks of spatial information. As argued in [4], most computational aspects of geographic information do not consider these linguistic intricacies, especially in the construction of formal models of spatial relations based on the logic of human spatial cognition. The authors also argue that such formal models do not acknowledge the way that people actually express spatial relations linguistically. Therefore, Bateman [3] suggests a semantics and linguistics based spatial ontology that would facilitate better mappings between spatial calculi and NL spatial expressions, such as the Generalized Upper Model ontology (GUM) [2].

In more recent works, researchers have been working on parsers that can process unrestricted language input, and thus, come closer to the parser discussed in this paper. For example, Zlatev [26] worked on the cognitive linguistic research in spatial semantics and provided basic theoretical concepts for understanding it, in the form of the holistic spatial semantic theory. Based on that theory, Kordkamshidi et al. [13] then implemented spatial role labeling (SpRL), a method with which they assigned reference objects, locata and spatial relations roles to linguistic terms. This technique facilitates mapping the terms onto formal spatial relations and is also used in this work. The authors suggest machine learning techniques to deal with ambiguity in linguistic spatial information and in [14], they utilize the proposition project (TPP) from SemEval-2007 [18] for disambiguating the spatial meanings of prepositions and enhancing the SpRL technique. Zlatev [26] worked on the cognitive linguistic research in spatial semantics and provided basic theoretical concepts for understanding it. Ever since then, there has been growing trend to understand natural language (NL) descriptions. The research has moved towards understanding unrestricted NL in current works ([14, 20, 11]). The methodology suggested in this paper makes use of no external resources (such as maps, vision), but employs a machine learning tool [19] to identify place names and the associated prepositional phrases. Hereafter, we work with the concept and reasoning approach to form rules for extracting useful spatial information from the text.

*3.1.1 Spatial Triplets*

A spatial triplet is a triple of a locatum (LOC), reference object (RO) and a spatial relation (r ). Here, RO is the object by which the location of the LOC is defined, using the prepositional relation r. For example, <Melbourne Hospital diagonally across Peter Doherty Institute> references Melbourne Hospital as the LOC in terms of Peter Doherty Institute as RO by the relation diagonally across as *r*. Vasardani et al. [24] provide more information on the theory and the structure of triplets. This method of defining spatial information is derived from the technique of Spatial Role Labelling (SpRL), which first appeared in [14], and which was also conceptually verified in Kracht *et al*. [\*\*] for many languages, as the *configuration* and *mode* layers of locative expressions. In SpRL…..[the rest of the text as is from now on].