The Role of Model Testing in Standards Development: The Case of ISO-Space

James Pustejovsky and Jessica L. Moszkowicz

Laboratory for Linguistics and Computation
Department of Computer Science
Brandeis University, Waltham, MA 02454
{jamesp, jlittman}@cs.brandeis.edu

Abstract

In this paper, we describe the methodology being used to develop certain aspects of ISO-Space, an annotation language for encoding spatial and spatiotemporal information as expressed in natural language text. After reviewing the requirements of a specification for capturing such knowledge from linguistic descriptions, we describe how ISO-Space has developed to meet the needs of the specification. ISO-Space is an emerging resource that is being developed in the context of an iterative effort to test the specification model with annotation, a methodology called MAMA (Model-Annotate-Model-Annotate) (Pustejovsky and Stubbs, 2012). We describe the genres of text that are being used in a pilot annotation study, in order to both refine and enrich the specification language by way of crowd sourcing simple annotation tasks with Amazon's Mechanical Turk Service.

Keywords: Spatial Information, ISO-Space, Annotation Standards

1. Introduction

Human languages impose diverse linguistic constructions for expressing concepts of space, of spatially-anchored events, and of spatial configurations that relate in complex ways to the situations in which they are used. One area that deserves further development regarding the connection between natural language and formal representations of space is the automatic enrichment of textual data with spatial annotations. There is a growing demand for such annotated data, particularly in the context of language technologies and the semantic web. Textual data routinely make reference to spatial relations between objects, as well as objects moving through space over time. Hence, verbal subjective descriptions of spatial relations need to be translated into metrically meaningful positional information. A central research question currently hindering progress in interpreting textual data, however, is the lack of a clear separation of information that can be derived directly from linguistic interpretation alone and that requiring contextually derived interpretations.

Because of these concerns, arriving at the appropriate level of detail for the interpretation of spatial expressions in language has proved to be a considerable challenge. Early work on the annotation framework, ISO-Space, in fact, followed the general strategy of specification design employed in the creation and development of TimeML and ISO-TimeML (Pustejovsky et al., 2005; Pustejovsky et al., 2010), where existing models of temporal relations were vetted and tested for linguistic adequacy as well as against an annotated corpus, i.e., TimeBank. The spatial domain, however, is considerably more complex than the temporal domain. For example, the spatial preposition *on* as in *The cup is on the table* or *The clock is on the wall* can have multiple interpretations while the temporal preposition *on* (e.g., *The party is on Wednesday*) is relatively easy to interpret. An additional major difference between the temporal and spatial domains is the degree of underspecification provided by existing spatial calculi, when compared to the meanings of spatial relations as used in language. In other words, the mapping from spatial expressions to identifiable values of spatial relations in a logic is not nearly as direct as that encountered in temporal annotation. Hence, the role of qualitative spatial relations (cf. (Randell et al., 1992; Kurata and Egenhofer, 2007)) may be somewhat limited in a specification for spatial expressions in natural language.

The point to be made here is that, over a new domain of analysis, such as the appropriate interpretation of spatial expressions, it is not obvious what values should be assumed for labels over the data. Unlike annotation tasks that can adopt existing labels and word senses (such as those provided by WordNet), some tasks require several rounds of modeling and annotating to account for their complexity adequately.

In this paper, we discuss the ongoing development of ISO-Space, which aims to be just such a specification. ISO-Space incorporates the annotations of static spatial information, borrowing from the SpatialML scheme (Mani et al., 2010), along with the annotation of movement and the location of events (Pustejovsky and Moszkowicz, 2008; Pustejovsky et al., 2011). This work is being conducted within the ISO TC37/SC4 technical subcommittee on language resource management as part six of the Semantic Annotation Framework, where the goal is to create a new standard for capturing spatial and spatiotemporal information.

We first describe the overall goals of the ISO-Space framework. This is followed by a general discussion of the annotation development cycle that drives the development of the specification towards those goals. The initial model of ISO-Space is then presented, followed by a description of how crowd sourcing with Amazon's Mechanical Turk service is being used to test and refine that model.

2. The Goals of ISO-Space

As discussed in (Pustejovsky et al., 2011), we assume the ISO CD 24612 proposed standard, where a fundamental distinction is made between *annotation* and *representation*

(Ide and Romary, 2004). To this end, ISO-Space makes a distinction between an *abstract syntax* and a *concrete syntax*, where the concrete syntax is exemplified by a particular XML encoding, and the abstract syntax defines the model and the structures constituting the information about space, directions, and movement that may be contained in annotations. The abstract syntax consists of a *conceptual inventory* (Bunt and Pustejovsky, 2010) and a set of syntactic rules defining the combinations of these elements. The conceptual inventory for spatial language annotation will minimally contain the following notions: locations; topological relations; directions and orientations; motion; paths, frames of reference;and time and space measures. We also assume a traditional linguistic classification of spatial expressions, identifying at least four grammatically defined classes: spatial prepositions and particles; movement and position verbs; spatial attributes; and spatial nominals.

In addition to capturing implicit spatial information, ISO-Space includes additional properties of locations such as orientation and metric relations between objects, the shape of an object, the size of an object, elevation, granularity, aggregates and distributed objects, and objects in motion. While a major focus of the ISO-Space effort is to encode as complete a range of verbal descriptions of spatial properties as possible, we will not discuss further properties of the specification here. Rather, we focus on the methodology adopted to identify the specific values associated with the spatial relations and their arguments. As we will see, it is with this task that a somewhat different approach to specification development is needed.

3. The Annotation Development Cycle

The development of the ISO-Space specification follows the MATTER cycle as described in (Pustejovsky, 2006; Pustejovsky and Stubbs, 2012). Following that strategy, we aim to look frequently at real text and adjust the specification of ISO-Space accordingly after several rounds of annotation. The MATTER cycle involves iterating over the following process sequence: *Model-Annotate-Train-Test-Evaluate-Revise*.

The "Model Testing" phase of this cycle (shown in Figure 1) involves iterating over model development followed by subsequent testing by annotation. Because this results in a sequence of Model-Annotation pairs, we call this the MAMA (or babbling) methodology. This Model-Annotate-(Model-Annotate)* technique assumes a classic iterative software development cycle, as applied to the creation of a rich specification language to be used for annotation.
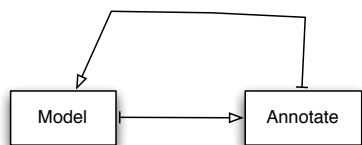

Figure 1: Model-Annotate Development

The MATTER and MAMA methodology represent a general strategy for standards development, particularly when the standard must account for relatively complex phenomena in natural language.

4. The Initial Model for ISO-Space

Initially we used the same technique that we had adopted for TimeML, creating tags from concepts found within the spatial semantics and qualitative spatial reasoning (QSR) literature that have served those communities well. The result was the early specification for ISO-Space using the following elements (Pustejovsky et al., 2011):

(1) a. ENTITIES: location, spatial_entity, motion, state, event_path, path;
b. SPATIAL RELATIONS: topological, orientational, metric.

For regions and geolocations, ISO-Space adopts the spatial location tag from SpatialML, called the PLACE tag, along with its attributes. ISO-Space also identifies the locations of movements and eventualities. For this, it assumes aspects of the Spatiotemporal Markup Language (STML) (Pustejovsky and Moszkowicz, 2008) and ISO-TimeML, to capture the spatio-temporal dimension. This is done by identifying spatial events involving motion as well as static situations. While identifying instances of motion is a fairly easy annotation task, when specifying spatial relations between objects, it can be quite difficult to determine which values to use for annotation. Adopting work from the QSR community, ISO-Space introduces a relation tag called a QSLINK ("qualitative spatial link"), which allowed different relation types, capturing the three types described in (1b). One of these, the *topological* QSLINK was specified as taking a value from the extended RCC8 set which includes RCC8 relations such as EC (touching), DC (disconnected), and PO (partially overlapping), as well as the IN relation introduced by SpatialML which is a disjunction of the RCC8 relations that imply that one region is contained within the other (TPP, NTPP, or EQ). However, unlike the fairly well-defined list of 13 values for temporal relations in language (as encoded in ISO-TimeML, for example), spatial prepositions are notoriously ambiguous and context dependent. Not only are there vastly more configurations possible between objects construed as spatial regions, but languages are idiosyncratic in how spatial information is encoded through different linguistic expressions. For this reason, we will have to define constraints that allow for underspecified semantic interpretations for several of the concepts introduced in our abstract syntax. These will need to communicate with various lexical (Fellbaum, 1998; Kipper et al., 2006) and spatial ontological resources (Bateman et al., 2010), in order to help disambiguate and more fully determine the semantics of relation types from the specification.

Because ISO-Space aims to account for a wide range of spatial language phenomena, the examination over a diverse set of corpora is crucial to account for adequate coverage of the phenomena. Using this initial ISO-Space specification, five separate genres were studied, where the specification was vetted and modified by a diverse working group. The chosen genres were written directions, standard newswire, location descriptions, interior descriptions, and travel blogs. Members of the working group examined multiple selections from each genre with an eye towards annotating with and improving the current specification. What

emerged from these meetings and the experiments with annotating the selected corpora was both interesting and quite enlightening. Here we will focus on the annotation of spatial relations.

The development of a reliable set of values for the labels over the data is a challenging task. Traditionally, data is labeled from an inventory of tags, each of which has a finite set of values associated with it. For example, a part of speech tag takes as its value something from the associated tagset, e.g., NN, VB, etc. Similarly, as mentioned above, in the context of labeling the temporal relation between two events in a text, the possible values are limited to (at most) the 13 relations from interval temporal logic (Allen, 1984). For simple geolocation relations as annotated in SpatialML, the somewhat generalized RCC relations are mostly adequate for expressing basic geo-topological relations (e.g., Rome is in Italy). However, when the spatial relation involved objects that were not identifiable spatial entities (i.e., LOCATIONS), the annotation became extremely difficult, as well as inadequate. As mentioned above, the values for the QSLINK involved a fairly direct mapping from a set of well-established spatial relations, as identified and developed within the qualitative spatial reasoning community, i.e., the RCC8 set. But several difficulties emerged when this finite set of relations was used to annotate spatial configurations between objects when with the use of spatial prepositions (e.g., *on*, *above*, *in*). The problem was that the values provided by these calculi were too restrictive or vague, where spatial configurations from natural language examples found in corpora simply were not distinguished by the relations provided by these calculi.

As an example, consider the limitations of the relation set from RCC8 as employed in distinguishing the senses of *on*, in the sentences below.

(2) a. There is a black stain *on* your shirt sleeve.
   b. The clock *on* the wall has the wrong time.
   c. The time *on* the clock reads 3:15 pm.
   d. Mary put the cup *on* the table.

The point here is not that different word senses for a preposition cannot be adequately distinguished. Rather, it is that often the sense distinction seems to be coming from a domain different from the one being used to annotate the expressions with, in this case, spatial features and relations. The question that arises is this: what is the utility of annotating a relation when the value supplied is so underspecified that it provides little or no information for subsequent inferencing (or question answering or translation)? Note that we are unable to use word senses as the values, since they are not available for prepositions (with the exception of PrepNet (Saint-Dizier, 2005; Saint-Dizier, 2006), which is not appropriate here). This problem presented itself with every spatial preposition, when used with non-geolocation entities. Clearly, we need to be able to provide a useful value to the spatial relation between objects, and the extended RCC8 values were not adequate to the task. In fact, topological relationships alone seem to be insufficient for distinguishing the different senses of spatial prepositions that occur regularly in natural language.

5. Testing the Model with Annotations

As a result of these experiments and the inadequacies of the initial model for relations, we adopted a different strategy for developing some aspects of the specification, using the MAMA methodology described above. The ISO-SpaceBank Corpus is an example of the development of the ISO-Space specification in this initial phase of the MATTER cycle. Through several iterations of small annotation tasks, the specification is being refined to the point that the remaining parts of the MATTER cycle can go forward. In addition, the resulting annotated text can be used as part of an ISO-Space corpus with little modification.

One of the key elements in the development of ISO-SpaceBank is to maintain a task-based methodology for its creation. That is, rather than attempting to annotate large swaths of text according to the current ISO-Space specification, small annotation tasks are defined that each address specific issues for the ISO-Space specification. These tasks include location/region identification, motion verb identification, spatial relation identification, the disambiguation of relation senses, spatial relation role labeling, and frame of reference labeling. Tasks are designed to be simple enough to allow for non-expert annotators to accomplish them; in this case, Amazon's Mechanical Turk (MTurk) service is being employed.

To begin, inherent locations and places are pre-annotated using gazetteers and geoname references. Given that as input, several rounds of MTurk annotation are performed that will eventually lead to an annotated corpus of spatial configurations. Specifically, tasks deal with spatial prepositions, static placement verbs, and motion verbs.

Presently, the spatial preposition MTurk tasks are the first to be run. To prepare the data for the MTurk annotators (referred to as "Turkers"), we began with the complete set of English prepositions before narrowing the focus to just those that have the potential to be spatial in some context. The set was then further reduced to the 25 most frequent potentially spatial prepositions and a corpus with about 100 uses of each preposition was created. The source of the data was the Berlitz Travel Guides section of the Open American National Corpus.

The first round of tasks using this data was to disambiguate the sentences to identify spatial uses. The Turkers were presented with a sentence with the preposition highlighted. They were then asked if the preposition as it was used in the sentence was spatial on non-spatial. Given these results, the second round of preposition tasks involved spatial preposition sense disambiguation using the Corpus Pattern Analysis (CPA) (Hanks and Pustejovsky, 2005; Rumshisky, 2011) technique. For this round, the Turkers are presented with a randomly selected target sentence using a given spatial use of a specific preposition. They are then asked to examine a series of sentences that use the same preposition and decide if the sense of the preposition matches that of the one used in the target sentence. Through several rounds of sense clustering, signature features for each sense can be identified.

The remaining MTurk task will proceed as follows. Round three of the spatial preposition tasks will identify the FIGURE and GROUND within the spatial preposition relation.

For static locational verbs such as *stand, sit*, and *touch*, a similar set of three rounds will be run. For motion verbs, the same strategy will again be employed with the addition of other arguments such as PATH, GOAL, and SOURCE. Looking ahead, additional crowd sourcing appropriate tasks for ISO-Space annotation will be identified.

6. Conclusion

In this paper, we described the need for an iterative development process when designing and applying annotation specifications. In particular, we discussed the ongoing development of ISO-Space, a language for the robust annotation of spatial information in natural language, and how a strict methodology of "Model and Annotate" (the MAMA method) can reveal gaps and uncertainties in a specification language, before it is broadened into a platform for wider use in annotation and adoption by the community.

In contrast to the MAMA strategy employed in the development of the ISO-Space specification, (Kordjamshidi et al., 2011) have approached the interpretation of spatial language in terms of pre-existing qualitative spatial relations. Their goal is to apply machine learning techniques so as to map the semantics of spatial language to qualitative spatial representations such as those in the Region Connection Calculus. This is a compelling approach to the problem of spatial language in text, but, as discussed earlier, it runs the risk of overlooking much of the complexity that is brought to bear when spatial relations are used in natural language. The MAMA strategy, adopted independently in (Müller et al., 2011) as well, should reveal much of that complexity, and it is expected that basic qualitative spatial relations will be insufficient to capture the full meanings of spatial relations successfully.

7. References

James Allen. 1984. Towards a general theory of action and time. *Arificial Intelligence*, 23:123–154.

John A. Bateman, Joana Hois, Robert Ross, and Thora Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence*.

Harry Bunt and J. Pustejovsky. 2010. Annotating temporal and event quantification. In *Proceedings of 5th ISA Workshop*.

C. Fellbaum. 1998. *WordNet: An electronic lexical database*. The MIT press.

P. Hanks and J. Pustejovsky. 2005. A pattern dictionary for natural language processing. *Revue Française de linguistique appliquée*, 10(2):63–82.

N. Ide and L. Romary. 2004. International standard for a linguistic annotation framework. *Natural Language Engineering*, 10(3-4):211–225.

Kara Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extensive classifications of english verbs. In *Proceedings of the 12th EURALEX International Congress*, Turin, Italy.

P. Kordjamshidi, J. Hois, M. van Otterlo, and M.F. Moens. 2011. Machine learning for interpretation of spatial natural language in terms of qsr. In *COSIT11*, pages 1–5.

Yohei Kurata and Max Egenhofer. 2007. The 9+ intersection for topological relations between a directed line segment and a region. In B. Gottfried, editor, *Workshop on Behaviour and Monitoring Interpretation*, pages 62–76, Germany, September.

Inderjeet Mani, Christy Doran, Dave Harris, Janet Hitzeman, Rob Quimby, Justin Richer, Ben Wellner, Scott Mardis, and Seamus Clancy. 2010. Spatialml: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44:263–280. 10.1007/s10579-010-9121-0.

A. Müller, C. Roch, T. Stadtfeld, and T. Kiss. 2011. Annotating spatial interpretations of german prepositions. In *Proceedings of Semantic Annotation Workshop ICSC 2011*.

James Pustejovsky and Jessica L. Moszkowicz. 2008. Integrating motion predicate classes with spatial and temporal annotations. In *Proceedings of COLING 2008*, Manchester, UK.

James Pustejovsky and Amber Stubbs. 2012. *Natural Language Annotation for Machine Learning*. O'Reilly Publishers.

James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39:123–164, May.

James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. Iso-timeml: An international standard for semantic annotation. In *LREC*.

James Pustejovsky, Jessica L. Moszkowicz, and Marc Verhagen. 2011. Iso-space: The annotation of spatial information in language. In *Proceedings of ISA-6: ACL-ISO International Workshop on Semantic Annotation*, Oxford, England, January.

James Pustejovsky. 2006. Unifying linguistic annotations: A timeml case study. In *Proceedings of TSD 2006*, Brno, Czech Republic.

David Randell, Zhan Cui, and Anthony Cohn. 1992. A spatial logic based on regions and connections. In Morgan Kaufmann, editor, *Proceedings of the 3rd Internation Conference on Knowledge Representation and Reasoningasoning*, pages 165–176, San Mateo.

Anna Rumshisky. 2011. Crowdsourcing word sense definition. In *5th Linguistic Annotation Workshop (LAWV) ACL-HLT2011*, Portland, Oregon.

P. Saint-Dizier. 2005. Prepnet: A framework for describing prepositions: Preliminary investigation results. In *Proc. of IWCS*, volume 6.

P. Saint-Dizier. 2006. Prepnet: a multilingual lexical description of prepositions. *Proc. of LREC 06, Genoa, Italy*, pages 877–885.