# Joint ISA-7 Workshop on Interoperable Semantic Annotation
# SRSL-3 Workshop on Semantic Representation for Spoken Language
# I2MRT Workshop on Multimodal Resources and Tools

## Workshop Programme

08.30 – 08:30 Registration
08:45 - 09:00 Workshop Opening

09:00 - 10:30 **Session: Semantic representation and multimodal resources**
09:00 - 09:30 Mehdi Manshadi and James Allen: *A Universal Representation for Shallow and Deep Semantics*
09:30 - 10:00 Rodolfo Delmonte and Agata Rotondi: *Treebanks of Logical Forms: They are Useful Only if Consistent*
10:00 - 10:30 Hennie Brugman and Mark Lindeman: *A Publication Platform for Open Annotations*

10:30 - 11:00 coffee break

11:00 - 13:00 **Session: Annotation of spatial information**
11:00 - 11:30 James Pustejovsky, Jessica Moszkowics and Marc Verhagen: *The Current Status of ISO-Space*
11:30 - 12:00 Robert Gaizauskas, Emma Barker, Ching-Lan Chang, Leon Derczynski, Michael Phiri and Chengzhi Peng: *Applying ISO-Space to Healthcare Facility Design Evaluation Reports*
12:00 - 12:30 Antje Müller: *Location and Path - Annotating Sense of the German Prepositions "auf" and "über"*
12:30 - 13:00 Linda Meini, Giovanna Marotta, Leonardo Lenci and Margherita Donati: *An XML Annotation Scheme for Space in an Italian Corpus*

13:00 - 14:00 lunch break

14:00 - 16:00 **Session: Semantic Roles and their annotation**
14:00 - 14:30 *Project ISO-Semantic Roles* (Martha Palmer)
14:30 - 15:00 Claire Bonial, Weston Feely, Jena Hwang and Martha Palmer: *Empirically Validating VerbNet using SemLink*
15:00 - 16:00 *The Lexlink project* (Collin Baker, Christiane Fellbaum, Martha Palmer)

16:00 - 16:30 tea break

16:30 - 18:00 **Session: Interoperable semantic annotation in ISO projects**
16:30 - 17:00 Kiyong Lee: *Interoperable Spatial and Temporal Annotation Schemes*
17:00 - 17:30 Harry Bunt, Rashmi Prasad and Aravind Joshi: *First Steps Towards an ISO Standard for Annotating Discourse Relations*
17:30 - 18:00 *Project ISO-Basics: Principles of Semantic Annotation* (Harry Bunt)
18:00 Workshop Closing

## Editors


| | |
|---|---|
| Harry Bunt | Tilburg University |
| Manuel Alcantara-Plá | Universidad Autónoma de Madrid |
| Peter Wittenburg | Max Planck Institute for Psycholinguistics, Nijmegen |


## Workshop Organizers/Organizing Committee

## Workshop Programme Committee

| | |
|---|---|
| Jan Alexandersson | DFKI, Saarbrücken |
| Stefan Baumann | Universität  Köln |
| Jonas Beskow | KTH, Stockholm |
| Paul Buitelaar | National University of Ireland, Galway |
| Harry Bunt | Tilburg University |
| Thierry Declerck | DFKI, Saarbrücken |
| Raquel Fernandez Rovira | Universiteit van Amsterdam |
| Anette Frank | Universität Heidelberg |
| Dafydd Giboon | Universität Bielefeld |
| Koiti Hasida | AIST, Tokyo |
| Nancy Ide | Vassar College, Poughkeepsie, NY |
| Michael Kipp | University of Applied Sciences, Augsburg |
| Kiyong Lee | Korea University, Seoul |
| Inderjeet Mani | Chiang Mai, Thailand |
| Jean-Clause Martin | LIMSI, Orsay |
| Lorenza Mondada | Université de Lyon 2 |
| Martha Palmer | University of Colorado, Boulder |
| Volha Petukhova | Vicomtech, San Sebastian |
| Andrei Popescu-Belis | Idiap, Martigny, Switzerland |
| Rarhmi Prasad | University of Wisconsin, Milwaukee |
| James Pustejovsky | Brandeis University, Wlatham, MA |
| Laurent Romary | INRIA/Humboldt Universität Berlin |
| Oliver Schreer | Fraunhofer Institute for Telecomuunications, Berlin |
| Mark Steedman | University of Edinburgh |
| Mariët Theune | Universiteit Twente |
| Isabel Trancoso | INESC, Lisbon |

# Table of contents

# Author Index

# Introduction

Three initiatives have joined forces in this workshop, which is concerned with issues in semantic annotation for language resources, especially in relation to spoken and multimodal language data, and with the interoperability and integration of resources and tools.

**ISA-7** is the Seventh Workshop on Interoperable Semantic Annotation, and forms part of a series of workshops of ISO TC 37/SC 4 (Language Resources) jointly with ACL-SIGSEM (Computational Semantics). These workshops bring together experts in the annotation of semantic information as expressed in text, speech, gestures, graphics, video, images, and in multiple modalities combined. Examples of semantic annotation include the markup of events, time, space, dialogue acts, discourse relations, and semantic roles, for which the ISO organization pursues the establishment of annotation standards, in order to support the creation of interoperable semantic resources.

**SRSL-3** is the Third Workshop on Semantic Representation of Spoken Language in Speech and Multimodal Corpora. In these workshops researchers convene who are working on speech and multimodal resources for the semantic annotation of related corpora, and take their inspiration from the observation that the semantic gap between the content conveyed by speech and other modalities and their formal representation is a burning issue in a range of tasks such as content mining, information extraction, dialogue processing, interactive story-telling, assisted health care,and human-robot interaction.

**I2MRT** (Integration and Interoperability for Multimodal Resources and Tools) is an initiative to address infrastructure aspects of the creation and use of interoperable multimodal resources. Main objectives of I2MRT are to create awareness of the need to make multimodal data visible via standardized methods and accessible via registered data centers; to discuss possibilities of harmonization and standardization of multimodal annotation schemes and possible mappings between schemes; to discuss ways to make cutting-edge technologies available to multimodality researchers; and to build a community that is committed to work further on these issues.

Harry Bunt
Manuel Alcantar-Plá
Peter Wittenburg

# Location and Path – Annotating Senses of the German Prepositions *auf* and *über*

**Antje Müller**
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
mueller@linguistics.rub.de

**Abstract**

Many difficulties concerning so-called spatial prepositions arise from an insufficient subclassification of the prepositions' interpretations. Since there is no one-to-one mapping from possible locations to prepositions there is a substantial need to differentiate the diverse interpretations of one preposition. In this paper we present an approach for a subclassification of some spatial prepositions. We will focus on the correlation between the German route prepositions *über* and *durch* and their static local counterparts *auf* and *in.* Route prepositions are often considered to be decomposable in a PATH function and a location. We will show that this assumption plus an adequate description of the underlying location results in a systematic classification of preposition senses. It is useful for the annotation of spatial preposition senses as well as for the analyses of the interpretations. For annotation the spatial interpretations are organized in a categorization tree. On the way through the tree different features are picked up that determine the respective interpretation. So every interpretation can be characterized as a set of features paired with the form of the preposition. This set-theoretic view of interpretations makes semantic relations between different interpretations of one and the same prepositions as well as between related interpretations of different prepositions apparent.

## 1. Introduction

The semantics of prepositions in actual usage reveals a high potential to express relations in various dimensions. Traditionally, the subcategorization of prepositions is based on apparently predominant senses of the respective preposition, so that prepositions are called spatial, temporal, modal, and causal (among others). But since prepositions show a huge diversity in their interpretations it is often not sufficient to simply call a preposition spatial or temporal.[1]

We adopt a perspective that we term a *relational* analysis of prepositions. A major feature of this view is that it is agnostic with respect to primary or prototypical meanings of prepositions. We assume that the prepositional system of a language is determined by mapping sometimes arbitrary subsets of relational meanings to preposition lexemes.

In this work we present a sense-based approach to some spatial prepositions with the focus on similarities between the so-called path prepositions and static locative prepositions.

The current approach is part of an ongoing study to develop a comprehensive as well as differentiated annotation scheme for preposition senses, starting with an analysis of 22 simple German prepositions.[2] From the subset of these 22 prepositions, we have identified those in (1) as prepositions that show a spatial interpretation. Other prepositions that may have prototypical spatial interpretations but are beyond that range will be left aside.

*(1) an, auf, bei, durch, gegen, hinter, in, mit, nach, neben, über, um, unter, vor*

Among these, we will focus on the prepositions *auf* ('on'), *in* ('in'), *über* ('above', 'over') and *durch* ('through'). We will show that a closer look at their senses combined with assumptions on the relation between prepositions referring to locations and paths can lead to a helpful classification of preposition senses in German.

The remaining paper is structured as follows: First, we will have a few words on a traditional classification of spatial prepositions and the relation between location and path in section 2. In section 3 we will focus on a sense differentiation of the preposition *auf* and the related path preposition *über*, and how their interaction fits the before mentioned assumptions. Then, we will present the results in the form of a classification tree for preposition senses as well as in a feature based representation related to it. The tree guides the annotator through the classification. Since every step in the tree corresponds to a feature immanent for the relevant interpretation of a preposition, in the end every interpretation is built from a set of features. This gives the opportunity to systematically account for semantic relations between interpretations in a set-theoretic view. The closer different interpretations are related to one another, the more features they share.

## 2. A system of spatial prepositions

### 2.1 Traditional classification

Traditionally, spatial preposition are subdivided into prepositions conveying simple localizations and path-related prepositions. Topological prepositions express relations between objects; projective prepositions carry information about a direction of these relations. This information makes projective prepositions sensitive to the frame of reference they are used in, whilst topological prepositions are independent from the reference frame established.

Path-related prepositions can be split into target (or goal) preposition, source prepositions, and path (or route) prepositions. In German, the form of most target prepositions is identical to the form of their static local counterparts. The distinction is correlated with a case alternation instead: Whereas prepositions expressing simple localizations govern the dative case, the target prepositions govern accusative case, as is illustrated in (2) and (3).

[1] See Müller et al. (2011) for an overview of other interpretations for 'prototypical' spatial prepositions.

[2] A scheme for the annotation of preposition senses for German prepositions did not exist prior to our investigation. The 22 prepositions under investigation are those simple prepositions of German that select NP as well as bare N complements and govern a case.

(2) *Die Katze liegt vor dem Bett.*
The cat lies in front of the bed.DAT
'The cat lies in front of the bed.'
(3) *Die Katze legt sich vor das Bett.*
The cat lies down in front of the bed.ACC
'The cat lies down in front of the bed.'

This alternation does not apply to every preposition. The preposition *bei* ('at', 'by', 'near'), e.g., never takes an accusative complement. In most cases the directional counterpart of *bei* is *zu* ('to').

As Kaufmann (1993) has pointed out, a path described by goal prepositions differs in his properties from a path described by path or route prepositions like *über* ('over', 'across') and *durch* ('through'). It is defined by a change of location to a given endpoint (defined by the reference object (RO)), starting at some point that can only be determined by not being this endpoint.[3] In virtue of this definition it has to be directed. Route prepositions, however, are used to describe continuous, not necessarily directed paths ((4), (5)). In the following, we will use the term path or route prepositions only for the route prepositions in the narrower sense. Target and source prepositions will be referred to just as such.

(4) *Sie geht stundenlang über die Wiesen um Blumen zu suchen.*
'She walks across/over the meadows for hours, searching for flowers.'
(5) *Die Straße führt durch den Ort.*
'The street goes through the city.'

As can be seen in (5), paths as such are non-temporal, even though many paths come together with movement and a change of location over time.

### 2.2 Location and path

The system of spatial prepositions is commonly assumed to be composed of some locative primes, depicted with labels like AT, IN or OVER, and the path-functions SOURCE, GOAL and ROUTE (Bennett, 1975; Jackendoff, 1983; Zwarts, 2005 among others) – see Table 1 for an example for English prepositions. Other prepositions are assumed to make static localizations at the regions referred to by the primes.

| | | SOURCE | GOAL | ROUTE |
|---|---|---|---|---|
| AT | *at* | *from* | *to* | *via* |
| IN | *in* | *out of* | *in(to)* | *through* |
| OVER | *on* | *off* | *on(to)* | *over, across?* |

Table 1: English adpositional system (Zwarts, 2010:13 adapted from Bennett, 1975:19)

It depends on the kind of path-function, whether a path must start at a given place, end at it, or intersect with it. For goal prepositions the end-point of the path is in a given place, for source prepositions the path has to start there. With regard to route prepositions, the given place must be intersected (or at least one intermediate point of the path has to be in this area). As the starting point for all kinds of these functions (and prepositions) remains the same, this should lead to a systematic deducibility of different interpretations of path prepositions from the interpretations of the locative prepositions in question (and vice versa). Once, the underlying region is defined we should get the interpretation of prepositions by simply applying the relevant function to it.

Our approach highlights the weak point of previous studies like Jackendoff (1983) or Bennett (1975): the insufficient definition of the location that serves as the starting point of the spatial functions. Others, like Wunderlich and Herweg (1991) try to define the relevant regions more precisely but only mention one possible interpretation of a preposition and with it only one possible locative starting point. Since there is no one-to-one mapping from regions to prepositions, there is a crucial need to differentiate between different senses of one and the same preposition and the regions connected with them. As Levinson and Meira (2003) pointed out (for topological prepositions), a direct coding of only a few locative primes (or near primitives) is not warranted cross-linguistically, as too much variation can be found.[4] Hence, it is required to define the spatial starting points for each language; and one cannot only refer to universal location primes. The regions one can point to with a preposition must be defined language dependent and on a more precise level. The relation between static locative prepositions and path prepositions can be helpful to verify assumptions about underlying locational concepts, since both should be traceable back to them. We will show that our approach to categorize spatial preposition senses supports such a decompositional analysis of (at least a subclass of) spatial prepositions.

## 3. Prepositions and spatial senses

In this section we like to present a differentiation of the spatial senses of *auf* and their connection with other static senses of spatial prepositions as well as with the path prepositions *über* and *durch*.

### 3.1 Classifying preposition senses

As was pointed out in Müller et al. (2011) our aim is not to classify spatial prepositions but the different spatial senses that are associated with the respective prepositions. We will illustrate this with the help of the preposition *auf*. We do not consider *auf* as a topological preposition but distinguish between a topological 'boundary area' interpretation (6) and a projective 'vertical' interpretation (7) of *auf*.

(6) *Die Schrift auf dem Schild*
'The lettering on the sign'
(7) *Die Tasse steht auf dem Tisch.*
'The cup stands on the table.'
(8) *Die Lampe hängt über dem Tisch.*
'The lamp hangs above the table.'
(9) *Die Tasse steht unter dem Tisch.*
'The cup stands under the table.'

There is no relevant axis or directional vector with the

[3] Kaufmann defines this change of state as a phase quantifier in accordance with Löbner (1990).

[4] Admittedly, there seem to be universal tendencies to organize adpositions around some attractor areas in a semantic space.

first sense, the second one, however, involves the vertical axis of the reference object. *Auf* differs from a projective use of *über* ('above') (8) in imposing some restrictions concerning contact or support between the object to be localized (LO) and the reference object (RO) (Schröder, 1986; Wunderlich & Herweg, 1991). In building the opposite with *unter* ('under', 'below') (9), *auf* just behaves like *über* with its projective interpretation.

Additionally, we assume a sense of *auf* that is classified together with a sense of *in* as localizing the LO within the RO. This interpretation can be exemplified by sentences like (10), where *auf* can be interchanged with *in* (11). One could argue that in those examples *auf* offers a localization in a boundary area of the RO or simply means 'higher than (+contact)'. Such an analysis, however, could not account for the similarity of *auf* and *in* in the examples (11) and (10) nor for the possibility of example (12). In this example the verb *vergraben* ('buried') impedes an interpretation in which the LO is located higher than the RO but only allows for an interpretation of the LO being buried somewhere within the area of the construction site.

*(10) Es befanden sich Schweizer Truppen auf deutschem Gebiet.*
'There were Swiss troops on German territory.'
*(11) Es befanden sich Schweizer Truppen in deutschem Gebiet.*
'There were Swiss troops in German territory.'
*(12) Es lag auf der Baustelle vergraben.*
'It was buried on the construction site.'

But what is the difference between (11) and (10)? Let us take a look at some more examples, where *auf* and *in* are interchangeable.

*(13) a) Sie stand auf der Wiese.*
'She was standing on the meadow.'
*b) Sie stand in der Wiese.*
'She was standing in the meadow.'
*(14) a) Es steht eine Bank auf dem Hof.*
'There is a bench on the yard.'
*b) Es steht eine Bank im Hof.*
'There is a bench in the yard.'
*(15) a) Hans steht auf der Straße.*
'Hans stands on the street.'
*b) Hans steht in der Straße.*
'Hans stands in the street.'

In (13) a) one will think about a meadow with short grass while in b) the meadow is high grown and the person in it is (almost) covered. If (14) b) is compared to (14) a), some (assumed) buildings (in fact three-dimensional fences) around the yard seem to be more crucial for the interpretation. With (15) we can find the same effects. So in all those examples the second sentence creates an impression of a three-dimensional room while in the first examples there is an emphasis on the plane surface. The relevant property for distinguishing the use of *in* and *auf* seems to be the dimensionality of the RO. If the RO is conceptualized having three dimensions, the preposition *in* is used. If there is a conceptualization with less than three dimensions, we need to use *auf* instead.

### 3.2 Senses of path prepositions

A simple observation concerning path prepositions is that whenever *auf* is used for a static localization, *über* is used for a traversal. As well, whenever *in* is used for a localization inside the RO, *durch* is used for a traversal through it.

*(16) Er liegt auf der Wiese.*
'He lies on the lawn.'
*(17) Er geht über die Wiese.*
'He walks across the lawn.'
*(18) Er liegt im Wald.*
'He lies in the woods.'
*(19) Er geht durch den Wald.*
'He walks through the woods.'

The interchangeability found for *in* and *auf* in some environments can be observed for *über* and *durch* in the same environments as well.

*(20) a) Sie geht über die Wiese.*
'She walks across the meadow.'
*b) Sie geht durch die Wiese.*
'She walks through the meadow.'
*(21) a) Er rennt über den Hof.*
'He runs across the yard.'
*b) Es rennt durch den Hof.*
'He runs through the yard.'
*(22) a) Hans läuft über die Straßen.*
'Hans walks over the streets.'
*b) Hans läuft durch die Straßen.*
'Hans walks through the streets.'

If PATH were a function applicable to the same place as the location function, one would expect for the interpretations of path prepositions to share properties with the interpretations of the static local prepositions. Two consequences follow from this assumption: First, it should be possible to apply the same methods for a classification, we found useful for the interpretations of *auf,* and get a suitable categorization for the interpretations of the path prepositions. Second, it should be possible to deal with the interchangeability of *über* and *durch* equal to the interchangeability of *auf* and *in*.

For *auf* we considered three different interpretations: the 'projective vertical' interpretation, the 'boundary area' interpretation and the 'inside of' interpretation. Examples are repeated below.

*(23) Die Tasse steht auf dem Tisch.*
'The cup stands on the table.'
*(24) Die Schrift auf dem Schild*
'The lettering on the sign'
*(25) Sie stand auf der Wiese.*
'She was standing on the meadow.'

For all three interpretations path-equivalents can indeed be found. For all three, the paths have to pass the area that also is described by the static localizations in (23) - (25).

*(26) Er schiebt die Tasse über den Tisch.*
'He pushes the mug over the table.'
*(27) Die Schrift zieht sich über das Schild.*
'The lettering stretches over the sign.'

(28) *Sie geht über die Wiese.*
‘She walks across the meadow.’

As already mentioned, the projective interpretations of *auf* and *über* differ in establishing, respectively prohibiting, contact between the LO and the RO. Considering examples like (29) and (30) or (26) and (31) it seems to be appropriate to not only assume *über* as a path equivalent for the interpretations of *auf* but also for the projective interpretation of *über*.

(29) *Er klettert über die Mauer.*
‘He climbs over the wall.’

(30) *Er springt über die Mauer.*
‘He jumps over the wall.’

(31) *Er wirft den Ball über den Tisch.*
‘He throws the ball over the table.’

It is still open for clarification, whether the dimensional conditions concerning *auf* and *in* can be demised to *über* and *durch* (assuming the ‘inside of’ interpretation of both). Examples like (32) and (33) suggest that it is not sufficient to only look at the dimensions of the RO. Both reference objects are likely to be conceptualized with only two dimensions.

(32) *Die schwere Wanne bricht durch die Decke.*
‘The heavy tub breaks through the ceiling.’

(33) *Peter geht durch die Tür.*
‘Peter walks through the door.’

For path prepositions there is a new dimension, added by the path, one has to take into consideration. While with *durch* the path always describes the third, missing dimension, with *über* (in its ‘inside of’ interpretation) the path runs along one of the given dimensions of the reference object. Accordingly, it is not the dimensionality of the reference object but the number of dimensions relevant to the whole spatial arrangement determining the choice of preposition.

# 4. The resulting classification

## 4.1 A classification tree for spatial prepositions

### 4.1.1 The excerpt discussed

The annotation is guided by a classification tree for the spatial preposition senses, the design of which is based on the observations discussed here.

A localization usually is defined as denoting a relation between the LO and the RO. We, however, will establish a localization as the relation between the LO, the RO, and the relevant neighboring region of the RO, which we will explicitly refer to as RO*. This region RO* is the foundation of every preposition sense considered in Figure 1, since it is the region something has to be localized in. RO* is always defined in relation to the RO. The adoption of such an explicit reference to a preposition independent neighboring region RO* allows us to capture the very relations between path prepositions, goal prepositions and simple localizations by establishing the relevant constraints for the region only once. For static localizations, the LO is localized in the region RO*, for goal prepositions the end of the path (starting somewhere else) has to be in this region[5], and for route prepositions there has to be at least one intermediate point of the path in the region RO*.

The axes in the tree, however, define the locational properties of the region RO* in relation to the RO, as well as other conditions required for the localization.

Note that for static localizations and goal prepositions – the directional counterparts of those prepositions – the form of the preposition remains the same. Only the case governed will change.

Those interpretations on the path *RO* is outside the RO; RO* is on reference axis of RO* are the so called projective interpretations of prepositions, whereas the other interpretations mentioned are considered topological. This categorization is one that only holds for the relation between RO* and RO but not for a path traversing or ending in the region RO*, since there is no condition about a path being placed on or parallel to one of the reference axes.

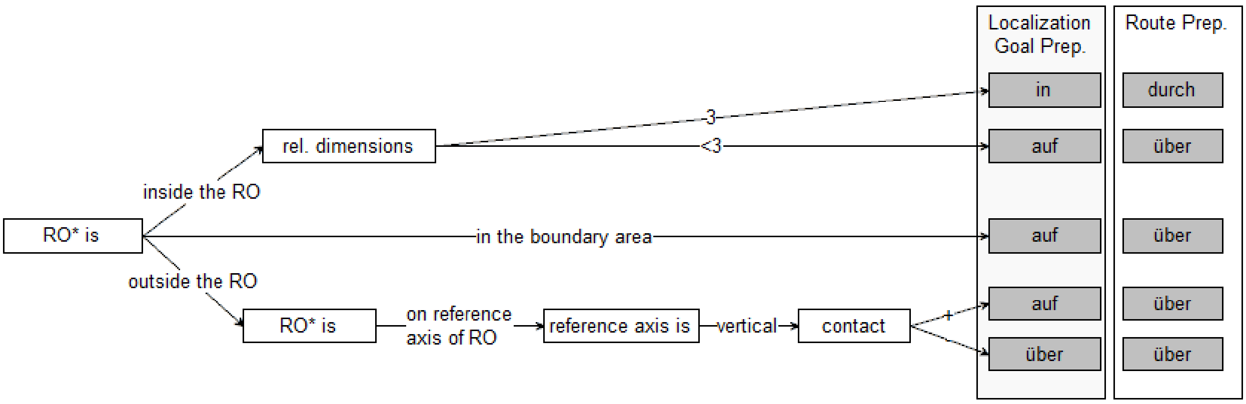


Figure 1: Excerpt of a classification tree for prepositions

[5] Or a CHANGE to this position has to be established.

#### 4.1.2. Other senses

Since we do not only want to consider the preposition senses explicitly discussed in here, and since we did not chose the prepositions under investigation by means of their systematic spatial relationships, a few problems for a classification tree and its implementation arise. As mentioned before, not every preposition has a directional counterpart only distinguished by case. As a result, we end up in investigating *bei* ('near', 'at') but not its directional pendent *zu* ('to'). What is more, not for every region RO* there is a path ending in it or crossing through it that can be described by a simple preposition. Take for example the region referred to by the preposition *hinter* ('behind'). A path passing this region has to be referred to by '*hinter…her'* as in (35).

(34) *Karl steht hinter dem Haus.*
'Karl is standing behind the house.'

(35) *Karl geht hinter dem Haus her.*
'Karl passes behind the house.'

As we only consider some simple prepositions, we do not account for situations like this. It should be noted, though, that an adding of senses and (complex) prepositions should be possible.

On these grounds, the impression of systematicity Figure 1 conveys cannot hold for the tree as a whole in Figure 2. There may be several alternatives to capture the facts mentioned. We decided to reduplicate some of the branches to account for exactly those path interpretations we needed for our set of prepositions. The local/directional differentiation, however, we account for with the feature [±DIR]. It will be added to a sense after the classification has been traversed (and is not applicable for the path prepositions) and is not listed as a separate feature in the figure. This possibility arises, since for all directional interpretations being considered, there is a static localization built with the same preposition. Furthermore we differentiate path prepositions in the narrower sense from other path related prepositions (not including path prepositions) since they impose different conditions on the paths. For a more detailed description of the senses not mentioned here see Müller et al. (2011).

### 4.2 A feature based representation

Corresponding to the tree, prepositions can be presented as a pair of their form and the features defining the interpretation. The interpretation is modeled from a set of features determining the preposition sense (*prep-sense*) and a set of constraints (*constraints*). The general form of a preposition can be given as follows:

*<<{Prep-sense},{constraints}>, prep-form>*

The division in *prep-sense* and *constraints* accounts for synonymy, polysemy, and antonymy relations. Polysemy is captured by different sets of *prep-sense* combined with the same *prep-form*.

Different *constraints* paired with the same set of *prep-sense*-features (and different *prep-forms*) result in near synonymy or antonymy. The more features preposition senses share, the more alike they are, the less they share, the less they have in common.

Consider the representations of the 'inside' sense of *auf* and *in*. They only differ in the constraint they impose on the relevant dimension ((36), (37)). The same holds for the projective interpretations of *auf* and *über*. They impose a different constraint on the setting, namely the existent or non-existent contact between LO and RO ((38), (39)).

(36) *<<spatial, {LOC(LO, RO*), inside(RO*, RO)},{relDIM=3}>, in>*

(37) *<< spatial, {LOC(LO, RO*), inside(RO*, RO)},{relDIM<3}>, auf>*

(38) *<< spatial, LOC(LO, RO*), outside(RO*, RO), vertical(RO*, RO)}, {contact(LO, RO)}>, auf>*

(39) *<< spatial, LOC(LO, RO*), outside(RO*, RO), vertical(RO*, RO), {-contact(LO, RO)}>, über>*

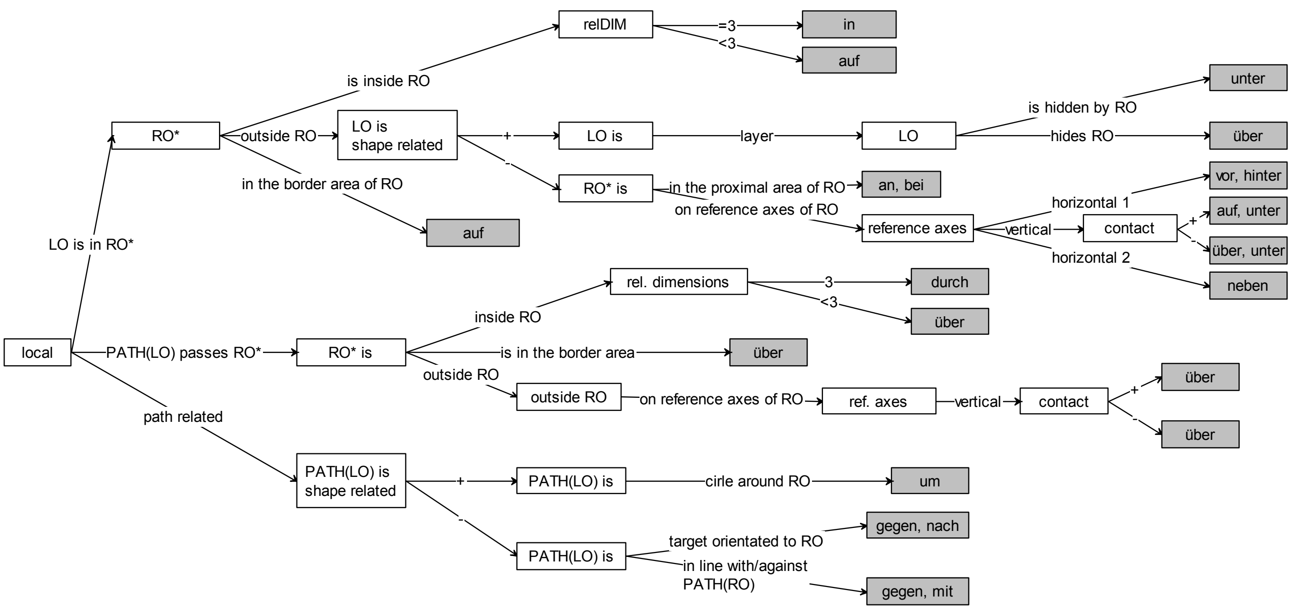


Figure 2: Classification of spatial prepositions

For an antonymic relation we can consider an interpretation of *über* and *unter* we have not mentioned yet.
For this interpretation an arrangement in layers is relevant, and *über* systematically interchanges with *unter* depending on one layer hiding the other or being hidden by it (40).[6]

(40) *Das Bild hängt über dem Loch.*
‘The picture hides the hole.’

(41) *Das Loch ist unter dem Bild.*
‘The hole is hidden by the picture.’

This correlation can be modeled by a constraint holding the inverse relation (in this case between the LO and the RO) for *über* and *unter*.

(42) *<<{spatial, LOC(LO, RO*), outside(RO*, RO), layer(LO)}, {hide(LO, RO)}>, über>*

(43) *<<{spatial, LOC(LO, RO*), outside(RO*, RO), layer(LO)}, {hide(RO, LO)} >, unter>*

The difference between path prepositions, goal prepositions and static localizations is captured by defining what is to be localized in the neighboring area of the RO.
We assume some kind of localization function LOC that has to be refined depending on the theory used. For static localization it simply takes the LO as an argument, for path prepositions we assume the argument has to be an intermediate part of a path, for goal prepositions the endpoint of the path.[7]
One has to be aware of the fact that this does not hold for all prepositions. There are prepositions that are always path related and impose other restrictions on the paths needed. The preposition *um* (‘around’) restricts the path to have a special shape, the target-orientated prepositions *nach* and *gegen* (‘towards’) demand for the endpoint of the path to be nearer to the RO than the starting point ((44), (45)).[8] The interpretation ‘in line with’ of *mit* (‘with’) (46) and its counterpart *gegen* (‘against’) (47) seems to need the RO to form a second path the first one can be orientated towards or against.

(44) *Das Pendel schlug nach der Seite aus.*
‘The pendulum swung to the side.’

(45) *Das Pendel schlug gegen eine Seite aus.*
‘The pendulum swung to the side.’

(46) *Ernst fotografiert mit dem Licht.*
‘Ernst takes a picture with the light.’

(47) *Ernst fotografiert gegen das Licht.*
‘Ernst takes a picture against the light.’

## 5. Conclusion

We presented a classification for a subclass of German prepositions with spatial interpretations. It is based on the assumption that there is no mapping from locative primes to prepositions but a need to define regions one can refer to with prepositions carefully. One preposition can refer to more than one region and therefore we need to distinguish different senses of prepositions. Relations between prepositions and their senses can help as clues for a finer definition of the relevant regions in one language. Additionally, they help to find other properties, like the shape of objects or different possibilities of conceptualization, relevant for a proper use of the prepositions.

## 6. Acknowledgements

We gratefully acknowledge the funding of our research by the *Deutsche Forschungsgemeinschaft* (DFG) under project grant KI 759/5-2.

## 7. References

Bennett, D.C. (1975). *Spatial and Temporal Uses of Prepositions: An essay in Stratificational Semantics*. London: Longman.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge MA: MIT Press.

Kaufmann, I. (1993). Semantic and Conceptual Aspects of the Preposition durch. In Zelinsky-Wibbelt, C. (Ed), *The Semantics* of *Prepositions*: *From Mental Processing* to *Natural Language Processing.* Berlin: Mouton de Gruyter, pp. 221--248.

Levinson, S., Meira, S, (2003). ‘Natural Concepts’ in the Spatial Topological Domain - Adpositional Meanings in Crosslinguistic Perspective: An Exercise in Semantic Typology. *Language* 79(3), pp. 485--516.

Löbner, S. (1990). *Wahr neben Falsch. Duale Operatoren als die Quantoren natürlicher Sprache.* Tübingen: Niemeyer.

Müller, A., Roch, C., Stadtfeld, T. and Kiss, T. (2011). Annotating Spatial Interpretations of German Prepositions. In O'Conner, Lisa (Ed.), *Fifth IEEE International Conference on Semantic Computing*. Stanford, CA, pp. 459--466.

Schröder, J. (1986). *Lexikon deutscher Präpositionen*. Leipzig: VEB Verlag Enzyklopädie.

Wunderlich, D., Herweg, M. (1991). Lokale und Direktionale. In A. von Stechow and D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin/New York: de Gruyter, pp. 758--785.

Zwarts, J. (2005). Prepositional Aspect and the Algebra of Paths. *Linguistics and Philosophy* 28.6, pp. 739--779.

Zwarts, J. (2010). A hierarchy of locations: Evidence from the encoding of direction in adpositions and cases. *Linguistics* 48.5, pp. 983--1009.

[6] See Müller et al. (2011) for a more detailed description of the sense.

[7] There is of course the possibility for goal prepositions to be defined with the help of a change function, taking the location function as an argument, just like illustrated in Kaufmann (1993).

[8] See Zwarts (2005) for more detailed remarks on restrictions for paths.

# Publishing and Exploiting Vocabularies using the OpenSKOS Repository Service

**Hennie Brugman [1], Mark Lindeman [2]**
1. Meertens Institute
P.O. Box 94264, 1090 GG Amsterdam
E-mail: hennie.brugman@meertens.knaw.nl

2. Pictura Database Publishing
De Hoefsmid 11, 1851 PZ Heiloo
E-mail: M.Lindeman@pictura-dp.nl

**Abstract**

Many vocabularies in eHumanities and eCulture domains can, and increasingly often are converted to SKOS. The OpenSKOS web service platform provides easy ways to publish, upload, update, harvest, query and distribute SKOS vocabulary data. This has benefits for vocabulary builders, vocabulary consumers and builders of tools that exploit vocabularies. In this paper we present and discuss the OpenSKOS system and a number of its applications, including an application from the domain of linguistic resources and tools.

## 1 Introduction

The application and relevance of vocabularies for the description of cultural heritage and scientific collections is making a comeback. One of the motivators for this comeback is the emergence of Semantic Web and Linked Open Data. There is much interest in application of data and text mining techniques to disclose collections, but it turns out that many of these techniques also build on vocabulary information.
Recent years have seen forms of standardization for vocabulary data that are consistent with Semantic Web and Linked Data principles. Well known is the W3C SKOS (Simple Knowledge Organization System) recommendation (Miles, 2009). More and more vocabularies, especially in the cultural heritage domain are mapped and converted to the RDF-based SKOS format and data model.

In 2004 the Dutch CATCH research programme started. CATCH (Continuous Access To Cultural Heritage) consists of a number of projects that do research regarding computer science and humanities research questions that are driven by cases from daily practice at large Dutch cultural heritage institutions. CATCHPlus is a partner project of CATCH that does valorization: it has the assignment to turn research prototype systems and demonstrators from the CATCH programme into tools and software services that can actually be used by cultural heritage professionals and users.

CATCHPlus tools and services should, where possible, contribute to the emerging infrastructure for digital cultural heritage. One aspect that many of the tools and services in CATCHPlus have in common is that they deal with or exploit vocabulary data. Therefore CATCHPlus stimulated standardisation of vocabulary formats to SKOS and also started work on a shared service that adds some standardisation to the way these SKOS vocabularies are made available and accessed: OpenSKOS[1], a web service based vocabulary publication platform.

Section 2 will describe requirements and motivations for OpenSKOS. Section 3 will describe the OpenSKOS architecture and components in detail, section 4 will position OpenSKOS in comparison with the ISOcat terminology service and with Linked Open Data. Section 5 describes current and future applications and clients of the OpenSKOS service. We will end the paper with an evaluation and conclusions (section 6).

## 2 Problem statement

The importance of and interest in vocabulary resources is increasing. These resources are typically created in specialized vocabulary maintenance tools or in modules of collection management systems. They are made available online using interactive web applications or in the form of Linked Data at the most. Over the last couple of years some standardization with respect to format has taken place: many vocabularies are currently mapped to SKOS.
However, it is often still a cumbersome process to locate suitable vocabularies and to (re)use them for one's own resource description tasks, in one's own tool environment. This is especially true when a vocabulary is well maintained and therefore frequently updated. To use a concept that is newly introduced by the vocabulary editors typically requires export and upload/download of the full vocabulary, proprietary format conversions and software adaptation or configuration steps by the producers of several collection management systems.

[1] http://openskos.org

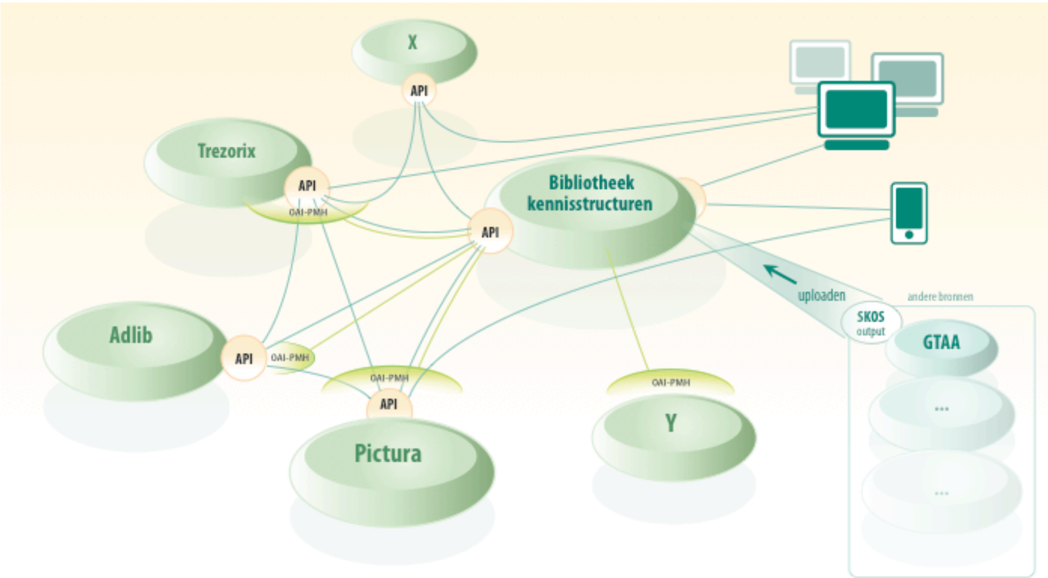


Figure 1: OpenSKOS architecture

Some web service based solutions also provide access to vocabularies as data, but these often have other shortcomings. They do not support periodic and/or incremental updates, they do not support the full underlying data model of the vocabularies (e.g. they are not able to handle relations between concepts), or they are optimized for other use cases than providing concepts for resource description (e.g. they have no proper support for handling long lists of entity names).

The Linked Data movement also imposes additional requirements on vocabulary services: concepts should be identified with stable, resolvable http URIs. Content negotiation is a desirable feature for a Vocabulary service.

Finally, web based (Open) Annotation (Sanderson, 2011) is a new development, that also imposes linked data type of requirements on Vocabulary services. It should be possible to annotate a web resource with URIs of concepts in online repositories.

## 3 The OpenSKOS service

OpenSKOS is a web service based approach to publication, management and use of vocabulary data that can be mapped to SKOS. The name is not meant to suggest that SKOS is not open; it refers to 'infrastructure and services to provide *open access* to SKOS data'. The main objective is to make it easy for vocabulary producers to publish their vocabularies and updates of it in such a way, that they become available to vocabulary users automatically and instantaneously, and independent of the specific software tools of these vocabulary users.

### 3.1 Architecture

Figure 1 shows the OpenSKOS architecture, which is a peer-to-peer architecture. Several sites can run instances of the freely available OpenSKOS repository software. Peers with a more centralized role are not technically necessary, although not excluded. Each site can be accessed by means of a RESTful API (Richardson, 2007) that supports a range of queries to retrieve or update SKOS vocabulary information in the repository. Having local copies of vocabularies in a repository instance implies that these can be searched efficiently on basis of locally created indexes.

Different OpenSKOS sites can exchange local copies of vocabularies using the OAI-PMH[2] protocol: OpenSKOS has built-in OAI-PMH data providers and harvesters. New vocabularies can be imported into the system in several ways: they can be harvested from another instance of OpenSKOS, they can be harvested from external OAI data providers, they can be included by implementation of the OpenSKOS API by other parties, or they can be uploaded using a built-in upload module.

Finally, OpenSKOS software contains a Dashboard to support a number of management tasks on each instance of OpenSKOS. This Dashboard can only be accessed after successful authentication.

### 3.2 The OpenSKOS RESTful API

The system's API is defined in a collaborative effort between the CATCHPlus project office, three major commercial tool providers for the Dutch Cultural Heritage sector (Adlib Systems, Pictura Database Publishing and Trezorix) and the Rijksdienst voor het Cultureel Erfgoed (Dutch department for cultural heritage). The specification is based on previous experiences and known use cases of all partners. The W3C SKOS recommendation was taken as the underlying data model.

#### 2.3.1 Functional scope of the API

To start with, the API can resolve (skos) Concepts and ConceptSchemes ('vocabularies') by URI in a number

[2]http://www.openarchives.org/OAI/openarchivesprotocol.html

of representation formats (JSON, RDF/XML, html). This implies that Linked Data access is a sub set of the web services functional scope. The resolve API has query parameters that allow filtering on language used, and specification of what information is/is not included in the result.
Second, the API has 'find' functionality for Concepts and ConceptSchemes. It supports a query parameter 'q' that takes queries according to the Apache Lucene Query Parser Syntax as values. Searching is possible over all SKOS based fields and over Dublin Core (dcterms) fields, if those are present. The result of a 'find' query is a list of Concepts (represented in the same way as for the concept resolve) and a diagnostics block, for example with number of results that match and number of results on page. Paging and sorting of results is supported.
A specialization of the /find API is the OpenSKOS 'auto complete' function, meant for interactive searching for matching concept labels starting with some characters. The primary use case for this auto complete is supporting resource description tasks in some collection or metadata management system.
The OpenSKOS API namespace contains *Collections* and *Institutions* that are not part of the SKOS model but added for practical reasons. Collections can group a number of conceptschemes together that constitute one resource from an organisational/data management perspective. For example, the thesaurus of the Netherlands Institute for Sound and Vision (archive of the Dutch public broadcast corporations) consists of six sub thesauri but is maintained and published as a whole. *Institutions* are added to make information available on the vocabulary publishers themselves, and to associate authorized vocabulary managers with.
The API explicitly covers SKOS properties that are used to define mappings between concepts, also mappings between concepts belonging to different conceptschemes. The OpenSKOS repository is also a place where mappings across vocabularies can be maintained and exploited.
The OpenSKOS API not only supports HTTP GET operations on the resources described before, but for many of those resources it also supports PUT, POST and DELETE operations. It is therefore possible to perform vocabulary maintenance tasks directly on the repository using the API. For REST examples see openskos.org.

The CATCHPlus project office and Pictura together have built an OpenSKOS implementation that includes an implementation of the API. This implementation is internally based on Apache SOLR. It also includes implementations of other OpenSKOS components: a Dashboard, OAI harvester and data provider (including a job scheduler) and upload module for SKOS uploads.

## 3.3 OAI-PMH and upload modules

There are in principle three ways to enter vocabulary data into the OpenSKOS repository: create it from scratch using the APIs PUT and POST operations, upload it using the built-in upload module or harvest it using the built-in OAI-PMH harvester and job scheduler.
OpenSKOS repositories are able to harvest vocabulary data or to provide harvesting access to specific vocabularies from other OpenSKOS instances. This harvesting can be done periodically and incrementally. OpenSKOS includes a job scheduler that can be configured to run periodic harvesting jobs.
Reasons to harvest vocabularies to one's own OpenSKOS instance are: it can be used for an initial full download, and it subsequently keeps vocabulary information up to date. Another reason could be to maintain a copy for local indexing and searching. A reason to provide access for harvesting by others: most efficient, flexible and controlled way to allow downloads of potentially large data sets (http could lead to long download times and time outs).

OpenSKOS has a built-in upload module that can only be operated by authorized users using the system's Dashboard.

## 3.4 Dashboard

For management tasks by authorized users the system has an interactive Dashboard component. After successful authentication a user can access several panes. The "Manage institution" pane allows the user to enter and modify institution metadata, like name, contact information and website. "Manage collections" presents the user with an overview of available collections, and allows the user to create new ones. These collections are associated with the users' Institution. Each collection has associated metadata, like title, description, links to websites, and license information (preferably Open Database licences, of course). Also, for each collection it is possible to specify whether it is harvestable by other OpenSKOS instances and if the associated data is imported by upload or by OAI-PMH harvesting. In the latter case the OAI data providers' base URL can be specified.
Collections are the unit of 'upload' or 'maintenance', and can consist of data for several SKOS ConceptSchemes.
The "Manage users" pane gives an overview of existing users, their email addresses, their access rights (do they have writing access using the API, using the Dashboard or both) and their API key. It also supports creation of new users.
Finally, the "Manage jobs" pane gives an overview of scheduled and finished harvest and upload jobs.

Institution and collection info can not only be inspected and modified using the Dashboard; it is also available to anyone for inspection using the relevant API calls, represented as RDF/XML, JSON or html. The html representation makes it possible to browse over the repository content starting at an Institution, via its Collections and ConceptSchemes to representations of the Concepts themselves.

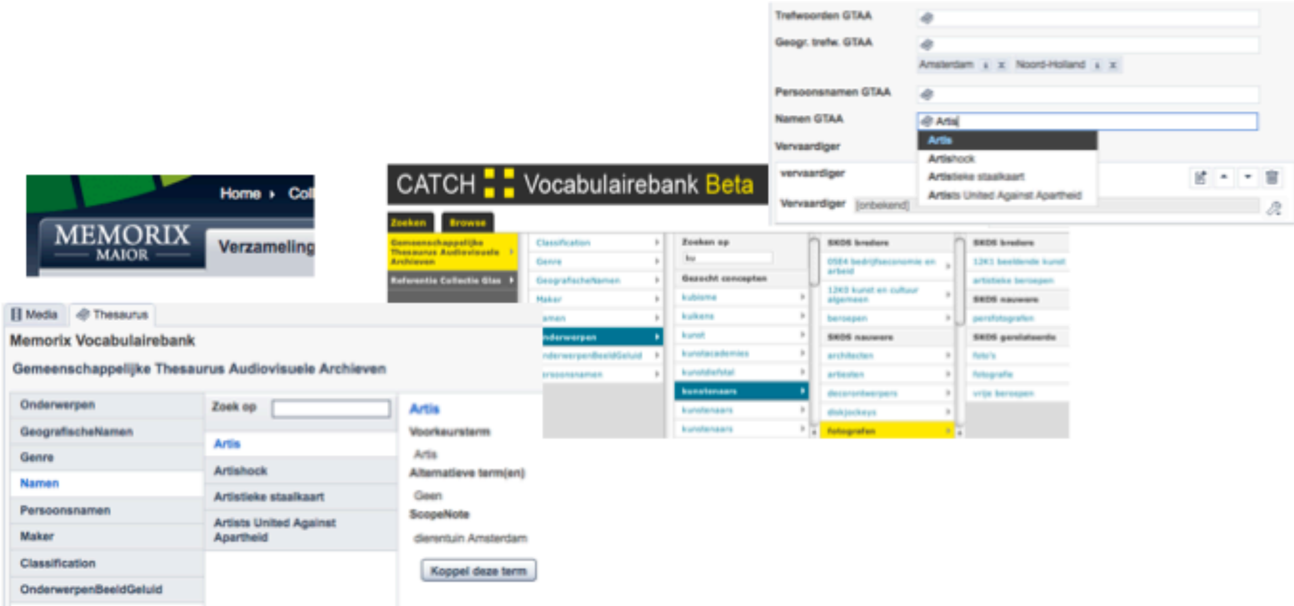


Figure 2: Snippets of user interfaces of OpenSKOS clients

### 3.5 Authentication and authorization

Since the main objective of OpenSKOS is to be 'open' we chose not to support authenticated 'read' access to the repository's content, all SKOS information is world-readable. In fact, we actively promote the use of open license forms like the Open Database license by offering this as an optional license form to creators of new vocabulary Collections.
For modification operations (create, update, delete) we support two levels of authorization: access using an API key, and access via the system's Dashboard. At API level modifications to Concepts and ConceptSchemes can be made. Modifications to Institutions, Collections and users all require authentication via the Dashboard.
Users can have either or both of the authorization levels.

## 4 Related work

OpenSKOS can in terms of genericity be positioned somewhere between a domain- and community-specific terminology repository solution as ISOcat and the generic and general purpose Linked Open Data approach.
ISOcat (Windhouwer, 2010) is an ISO TC 37 registry for Data Categories. These Data Categories are mainly intended for linguistic concepts. ISOcat by definition does not support relations between concepts and relies on separate relation registries for this. Main use cases for ISOcat are registration of concepts and providing a platform for standardisation of linguistic terminology. ISOcat therefore is not the optimal place to maintain or serve large lists of term labels. SKOS and OpenSKOS are less restrictive: they are not restricted to a certain domain, support relations between concepts and support a wider range of use cases. Representing and serving long term lists is normal practice. ISOcat has a RESTful web service that can be and actually is used to feed the OpenSKOS service (see chapter 5.3 about CLAVAS).
Linked Data on the other hand is even more generic: it is not restricted to vocabulary type of data, as SKOS and OpenSKOS are. It can represent any mix of data, metadata and concepts and links between those. The drawback is, that considered as a protocol it is much simpler than the ISOcat and OpenSKOS RESTful APIs. Linked Data access by means of resolvable and stable http URIs and support for content negotiation is a subset of the functionality of the OpenSKOS API.

## 5 Applications

The OpenSKOS repository service and architecture is the outcome of a process of several years, during which prototypes and experimental tools were built and tested. Over these years several academic, commercial and cultural heritage partners got involved. This section describes a bit of OpenSKOS' history and context, before it discusses current and planned applications of the system.

### 5.1 OpenSKOS history and context

Previous work in the CATCH research programme and in CATCHPlus resulted in a demonstrator and in a first version of the Vocabulary Repository service. This first version was implemented as a 'thin' Java layer on top of an RDF store (Openlink Virtuoso). Although stable and performant (e.g. online auto completion over the web works fine), this implementation makes a large demand on memory, and we had doubts about its scalability. Furthermore, its API is at best "REST-like", it has limited and incomplete support for modification operations, and there are no provisions for web upload, OAI-PMH harvesting or user authentication.
Nevertheless, this system was and is actually used for

daily collection description work by the triangle Netherlands Institute for Sound and Vision, National Archive, and Pictura and was found an elegant and interesting solution. (S&V is the thesaurus provider, National Archive does collection description with S&V terms using Pictura's Memorix tool).
This relative success led to intensive discussions between CATCHPlus, RCE, Adlib, Pictura, Trezorix that led to refinement of the OpenSKOS concept and a proper RESTful API specification that built on the knowledge, use cases and experience of all partners. Subsequently, the API, infrastructure and Dashboard were implemented by Pictura and CATCHPlus.
Due to this long history with frequent discussions, presentations and experiments in the Dutch cultural heritage context, there is now serious interest to participate. Several large Dutch CH institutions are currently involved in some way.
Recently CLARIN-NL also started a project to apply OpenSKOS for linguistic vocabulary data (see 5.3).

### 5.2 OpenSKOS clients

Some API clients already exist. A generic browse and search web application was built for CATCHPlus (by Q42, see figure 2). All access to vocabulary data used and shown in this web application is exclusively retrieved via API calls.
Pictura's collection management application Memorix is used on daily basis by National Archive for description of their online image collection. Memorix also functions as an OpenSKOS client.
Sound and Vision has started development of a web based thesaurus management application on top of the OpenSKOS editing APIs to manage their GTAA thesaurus.

### 5.3 Application by CLARIN(-NL): CLAVAS

Within the Dutch CLARIN context there turned out to be a need for an additional effort to promote uniform terminology. While ISOcat focuses on standardisation of sets of concepts (Datcats) there is an additional need for support of relative simple, but long lists of terms, especially in the context of metadata creation and editing. Therefore CLARIN-NL started the CLAVAS project, which is an application of OpenSKOS. The CLARIN project makes several contributions to OpenSKOS, and CLARIN in turn can benefit from additional efforts done for OpenSKOS. These contributions are three additional SKOS-ified resources (ISO 639-3 language codes, access to public parts of ISOcat through the OpenSKOS API and architecture, and a vocabulary of organisation names relevant for the international domain of linguistic tools and resources. It is explored if this list can be bootstrapped by existing metadata descriptions containing organisation information.
An additional CLAVAS component is a simple web application that supports basic vocabulary curation tasks on simple concept lists.
The CLAVAS project is done by the Meertens Institute, which also hosts the central CATCHPlus project office.

## 6 Evaluation and conclusions

The OpenSKOS service can be consulted in many use cases where vocabularies play a role. Some examples :

- When defining a metadata component, as for example in the CMDI framework it is possible to associate a metadata field with a ConceptScheme in OpenSKOS simply by associating the field with the URI of the ConceptScheme.
- When creating metadata in a metadata editor values for fields can be selected using the auto complete API of OpenSKOS.
- The service can be exploited in several browse in search scenarios, for example for faceted browsing or for query formulation.
- When Concepts have labels in multiple languages, localized views of metadata records can be displayed.

OpenSKOS supports all SKOS relations between Concepts, both within vocabularies and across vocabularies. SKOS and OpenSKOS also support enrichment of vocabulary concepts with links to other resources on the web (more specifically, in the Linked Data cloud).

Probably the greatest benefit of OpenSKOS is that it provides an easy publication platform for all resources that can be 'SKOS-ified'. This has advantages for vocabulary publishers, for vocabulary consumers and for builders of tools that create or exploit vocabularies.

Advantages for vocabulary publishers are:
- Offering vocabularies to others is as easy as a simple upload action.
- It is easy to use your own vocabulary in the tools of others, if these tools use OpenSKOS.
- Vocabularies can easily and frequently be updated without involvement of others.
- It is easy to link your own vocabulary to vocabularies of others.

Advantages for vocabulary consumers :
- Easy discovery, evaluation and reuse of existing vocabularies (and therefore a reduced need to construct your own).
- New browse and search possibilities.
- Always up to date versions of vocabularies are available

Advantages for tool builders :
- No more periodic updates, no more specific adaptations for specific vocabularies.
- Can benefit from efforts of other tool builders and of vocabulary publishers.

- Can use OpenSKOS API functionality for a range of use cases.

OpenSKOS is available as open source from GitHub, and as installable package. It is implemented on basis Apache SOLR technology in a scalable way. A community of OpenSKOS users is already emerging.

## 7 Acknowledgements

We would like to thank all people and institutions that contributed to the realization of OpenSKOS by investing time, energy and/or funding. We especially would like to mention RCE, Adlib Systems and Trezorix for their contributions to the definition of the OpenSKOS architecture and API, and the funders of CATCHPlus: the Netherlands Organisation for Scientific Research (NWO), and the Dutch ministeries for Education (OCW) and Economic Affairs.

## 8 References

Miles, A., Bechhofer, S. (2009). SKOS Simple Knowledge Organisation System Reference. *W3C Recommendation* 18 August 2009.

Richardson, L., Ruby, S. (2007). *RESTful Web Services: Web services for the real world.* O'Reilly Media. May 2007.

Sanderson, R., Van De Sompel, H. (2011). Open Annotation. Beta Data Model Guide. http://www.openannotation.org/spec/. 10 August 2011.

Windhouwer, M.A., Wright, S.E., Kemps-Snijders, M. Referencing ISOcat data categories. *In proceedings of the LRT standards workshop* (LREC 2010), Malta, May 18, 2010

# Applying ISO-Space to Healthcare Facility Design Evaluation Reports

**Robert Gaizauskas*, Emma Barker*, Ching-Lan Chang†, Leon Derczynski*,**
**Michael Phiri† and Chengzhi Peng†**

*Department of Computer Science, †School of Architecture
University of Sheffield
S10 2TN, UK
{initial.surname}@sheffield.ac.uk

**Abstract**

This paper describes preliminary work on the spatial annotation of textual reports about healthcare facility design to support the long-term goal linking of report content to a three-dimensional building model. Emerging semantic annotation standards enable formal description of multiple types of discourse information. In this instance, we investigate the application of a spatial semantic annotation standard at the building-interior level, where most prior applications have been at inter-city or street level. Working with a small corpus of design evaluation documents, we have begun to apply the ISO-Space specification to annotate spatial information in healthcare facility design evaluation reports. These reports present an opportunity to explore semantic annotation of spatial language in a novel situation. We describe our application scenario, report on the sorts of spatial language found in design evaluation reports, discuss issues arising when applying ISO-Space to building-level entities and propose possible extensions to ISO-Space to address the issues encountered.

## 1. Introduction

Identification and interpretation of spatial information in natural language is a topic of increasing interest in contemporary computational linguistics. Newly emerging techniques in language processing, based on standards for annotating spatial language such as SpatialML (Mani et al., 2010) and ISO-Space (Pustejovsky et al., 2011a), are capable of automatically identifying location references in text and grounding them, e.g via geo-coordinates as supplied in a gazetteer. This grounding information supports fusion of text accounts with other digital applications.

Following the construction or significant refurbishment of healthcare facilities, qualitative methods are often applied by healthcare and architecture professionals to gather evidence about which aspects of the design have worked and which have not. These evaluative studies form important knowledge resources for future similar projects during their inception. However, current practice in disseminating evaluation studies often amounts to no more than distributing a limited number of hard copies of lengthy reports. This effectively limits access to the content of the reports, leading to the findings of such studies rarely contributing to best practice.

To address this problem we have initiated an investigation into Annotated 3D Interactive Navigation (**A3DIN**), to radically enhance the accessibility and readability of the evaluation study documents. The end goal is to build a software prototype from a fusion of 3D virtual architectural modelling and spatial language processing, applied to a small scale case study, that will allow a user to navigate from a particular passage in a textual design evaluation report to an appropriate illustrative view within a 3D model and vice versa. As one of the first steps in this process, we have assembled a small corpus of design evaluation reports and attempted to annotate spatial entities and relations in a subset of these texts in accordance with the ISO-Space semantic annotation guidelines. To the best of our knowledge we are the first to apply ISO-Space at the building/sub-building scale – all previous applications appear to have been at the urban, interurban and geographic scale as referenced in text types such as newswire reports and cyclist blogs[1].

In this paper, we report on work in progress within the A3DIN project. Specifically, we: (1) discuss the A3DIN scenario in more detail and the pilot study we are engaged in to investigate it (Section 2.); (2) describe our small corpus of design evaluation reports and present examples of the sorts of spatial language that characterise these reports (Section 3.); (3) present and analyze issues arising in the application of ISO-Space to the reports (Section 4.); (4) discuss related work (Section 5.); and (5) conclude with a summary of lessons learned and suggestions for adapting ISO-Space to work in this domain.

## 2. The Application Scenario

Healthcare buildings represent a significant investment and their design directly influences the functions they support – poor design can have a severe negative impact. A design approach both geared towards the needs of end users (through e.g. Design Quality Indicators) and learning from past experience (Evidence-based design) is therefore valuable. Post-Occupancy Evaluations (**POE**s) are recognised as important during building commissioning and use. Information garnered from these is an essential part of the evidence-based design process as well as a means of determining if design targets of the building being evaluated have been met. Despite this, the commissioning of POEs in the past has not been routine. There is a considerable body of post-occupancy information available, however it is quite variable in the nature of its content.

Worldwide, undertaking of POEs varies. There are records of POEs in the US dating back to the 1960s (Preiser et al., 1988) and POEs have been carried out for many building

[1] See Section 5. below or, for example, the annotations in LDC corpus `LDC2008T03`. There is a suggestion in Pustejovsky et al. (2011a) that the ISO-Space working group is investigating interior descriptions with a view to improving the specification to address them, but we are not yet aware of any published outcome of this work.

types: offices, schools, courthouses, prisons, housing and so on. Healthcare POEs are less common. This may in part be due to the status of healthcare building commissioning – for example, many EU countries with widespread state-funded healthcare (such as Sweden and Denmark) have not had hospital building programmes until now. Although POEs can be undertaken at any point in a building's lifetime, the most common point of undertaking is early in the building's use.

In contrast, the UK has built almost two hundred under a public-private initiative in the past twenty years. This increased the requirement for feedback and for learnings from POEs to be taken into consideration for subsequent construction projects. POEs are now mandatory in certain areas. As a result, the Department of Health in the UK has not only developed detailed healthcare guidance and design/evaluation tools (e.g. ASPECT/AEDET; UK Department of Health (2008a, 2008b)) to improve design quality but has also been instrumental in seeking to draw out lessons learned via post-project evaluations incorporating POE.

The UK is a special case where there is an established healthcare construction programme and also sophisticated post-occupancy evaluation methods, and where buildings recently constructed under the program have been occupied long enough for these methods to be applicable. This generates a situation where there is both a new type of evaluation report and also, with more hospital building programmes starting in other countries, broad demand for the information contained in such reports.

Despite such sophisticated methods for carrying out evaluation studies, current practice in reporting these studies, as noted above, often takes the form of bound paper-based documents, of which only a limited number of hard copies are made available due to the production cost. The accessibility of these reports is therefore quite limited and they do not contribute to wider adoption of best practice as revealed by these studies.

From a usability perspective, these evaluation reports are problematic for a number of reasons. First, despite inclusion of images and fragments of floor plans, it is frequently difficult to properly interpret the text without "seeing" the aspect of design under discussion in the visual context of the building. Second, for a reader interested in a particular part of the building or aspect of the design, or wishing to quickly ascertain the positive or negative features highlighted by the report, detailed perusal of a lengthy document may be required. Given the move by architects to use 3D modelling tools in producing designs of buildings, it is natural to ask if building documentation, such as evaluation studies, could be linked to the 3D models so that readers could move between the visual and textual mediums to faciliate better understanding and more flexible access to information. For example, a reader could point to the part or aspect of the building of interest and be shown the portion(s) of the report discussing it; good or poor aspects of the design could be highlighted directly in the 3D model, e.g., by use of colours, to provide a visual summary of the report that users could interact with to access more specific information in the report.

Manually linking texts and 3D models is not feasible in general, and thus the linking process needs to be automated. To automate this process requires a number of technical capabilities which do not exist at present, or are only just beginning to emerge:

1. the ability to recognize references to places, spatially situated entities and spatial relations in text;
2. the ability to associate semantic information with graphical elements in CAD-generated 3D models;
3. the ability to interpret spatial language in text in order to:
   (a) ground spatial referring expressions in the coordinate system of the graphical model;
   (b) model spatial relations holding between spatial entities (e.g. *The waiting area is adjacent to the courtyard*);
   (c) understand the viewpoint taken in the text (e.g. *As you enter the building the reception desk is easy accessible* ...)

   so as to present the correct portion of the model at the correct orientation and scale.

ISO-Space is an important step on the path towards achieving capabilities 1. and 3. Capability 2. is outside the scope of this paper, but is being addressed within the building design community, particularly through Building Information Modeling (**BIM**)[2] and the emergence of open standards to support BIM, such the Industry Foundation Classes model[3] which is in the process of becoming an ISO standard and is now implemented in open source tools such as BIMServer[4]. For present purposes, the key observation is that we can safely assume there will be some mapping between natural language terms and labels attached to semantic elements within a building design model, such as for instance room numbers or (possibly ambiguous) names for specially designed spaces, such as *waiting room*, *pharmacy*, *physiotherapy gym*, etc. These elements within the model are in turn associated with specific parts of the graphical representation of the 3D model which is itself positionally specified in terms of offsets from national survey benchmark points, thus indirectly grounding the whole model in the conventional geospatial co-ordinate system. From these mappings can be distilled the equivalent of a building-specific gazetteer – a resource mapping linguistic references to places within a building to portions of a 3D model and to spatial areas within the world.

To engage with this scenario we have chosen to investigate the design of a specific health care facility, the Jordanthorpe Health Centre in Sheffield, UK (Figure 1). We are in the process of building a 3D model for the site using Graphisoft ArchiCAD BIM software[5], with models exported to IFC

[2] See http://en.wikipedia.org/wiki/Building_Information_Modeling

[3] See http://www.buildingsmart.com/

[4] See http://bimserver.org/.

[5] See http://www.graphisoft.com/products/archicad/.

Figure 1: The Jordanthorpe Health Centre

format files, and have a collection of design evaluation reports about it, written by Masters level students from the School of Architecture, University of Sheffield. We also have detailed floor plans for the facility and are creating a "building gazetteer" from these to serve as a temporary 2D grounding target for spatial language recognition while the 3D model is being developed (linking textual content to a 2D floor plan representation is itself a challenging and worthwhile goal, as some textual observations are better illustrated by a 2D view from above, than from a 3D view from within).

## 3. Spatial Language in Design Evaluation Reports

To investigate the sorts of language used in design evaluation reports we have assembled a small corpus of four reports in English written by students from the School of Architecture, University of Sheffield, as part of their professional training[6]. These reports are shorter than those typically created by practising professionals, but otherwise are entirely realistic as they are created using the guidance and instruments recommended by the UK Department of Health. The documents range from 18 to 38 pages including images, or about 3000-10,000 words. They follow a standard report format and contain mainly evaluative and descriptive statements concerning aspects of the building design. As such, they are a particularly rich source of spatial language. Key features we observed in these reports include the following.

### 3.1. Multiple Scales

The majority of sentences in the reports contain expressions which refer to locations and entities at or below the level of the building, describing both the interior and exterior spaces of the site, for example: entrance, corridor, building interior, wall, ceiling, waiting area, door, windows, car park etc. (see example (1)). However, we also find expressions relating to locations and entities above the building level, such as counties, regions, cities, streets etc. (see (2)). There are also examples, such as in (3), of references to astronomical bodies (e.g. the sun).

(1) *The main entrance to the building is located in a corner under an overhang, which does not allow it to be visible to patients easily.*

(2) *The Jordanthorpe area is situated in the Southwest of the city of Sheffield, close to the border with Derbyshire.*

(3) *The sun rises in the morning behind the centre and moves in the direction shown on the sun path diagram below ...*

### 3.2. Multiple Perspectives

We find various types of location expressions, which reflect different perspectives on a space. For example, we can distinguish between:

1. terms referring to concrete architectural elements, e.g. *building*, *rooms*, *main entrance*, *corridor*, *car-parking spaces*, *windows*, *façade*, etc. Such terms may indicate function, e.g. *consultation rooms*, *main reception*, *patient female WC*, and often correspond to names in the associated floor plan.

2. more abstract expressions referring to areas or zones. These are typically (but not necessarily) labelled according to the function of the space or the category of intended user, e.g. *waiting area*, *parking area*, *designated queueing area*, *patients activity zone*, *staff only zone*. Such references often correspond to labelled areas in the building plan and we find they refer to multiple or partial spaces as denoted by the kind of references we refer to in 1.

3. expressions in the texts where these different perspectives are mixed and presented in relation to each other. For example:

   (4) *... the waiting area on the first floor has a great view of the courtyard as well as the front yard and the woods nearby.*

   (5) *The immediate interior area around the entrance feels reassuring because it is open and airy.*

Thus we see a complex mix of formal vs. functional terms, viewed at varying levels of granularity.

### 3.3. Spatial Relations

The reports contain a particularly rich set of spatial relation expressions. These include expressions relating the positions of locations or spatial entities to each other:

(6) *For example, the bottom corner of the pillar to the right of the entrance has a small area where the render is missing.*

(7) *The entire health centre is surrounded by a fence approximately 2m in height.*

[6] Reports were conducted under the Module ARC6810 "Architecture and the Design Process" during 2009-10 and according to University of Sheffield Architectural Healthcare Environment Research Group standards.

and also expressions relating other spatial aspects of entities, such as their relative size or their distance from each other (sometimes including measures):

(8) *These buildings, shown photographed from across the car park in Figure 4.2, are substantially smaller than the health centre ...*

(9) *The bus stop is a very short walk (approximately 30m) from the main entrance.*

### 3.4. Direction, Orientation and Viewpoints

In addition to the spatial relations just discussed, there are frequent examples of expressions which indicate compass direction or orientation. This may be the orientation of a particular entity/location, e.g., (4) above and also:

(10) *The front of the centre, where the main entrance is situated, faces towards the west.*

(11) *Most of the windows in the consultation rooms overlook the courtyards*

We also find examples of references to entities or locations which can be viewed from a particular position, and possibly via another: (a) viewed from (b), via (c)

(12) *View from the waiting area towards reception, showing the mezzanine floor, which adds interest to the interior form.* (a photo caption)

(13) *For example, the entire south façade of the building is fully glazed. This provides a view of both the ground and sky, a key design feature.*

### 3.5. Movement of Entities in Space

While the design evaluation reports are very rich in references to locations, spatial entities and their spatial relations, there are relatively few descriptions of motion (it is, after all, a static entity that is being evaluated). However there are some. These tend to refer either to the movement of light or air, or to paths patients will follow in using the facility. In both case reference is not a specific event (of air entering or of a patient moving) but rather of regular occurrence of events of a particular type.

(14) *As cold air can easily penetrate through the windows ... it can affect occupants thermal comfort.*

(15) *It is the first area that visitors will arrive at when using the car park or nearby bus stop.*

### 3.6. intentional Contexts, Modality, Negation and Conditionality

In contrast to the expressions which indicate the intended or actual function of a space, we also find expressions which indicate expected or believed consequences of design decisions, possible future use, the absence of things in space and conditional expressions, or combinations of several of these (cf. examples (16) and (17)), particularly in evaluative passages in the reports when missing features or alternative possibilities overlooked in the design are being pointed out.

From a linguistic perspective we see intentional contexts:

(16) *It is expected that the lack of blinds available to exclude sunlight could cause discomfort to both patients and staff.*

(17) *The author suggests that the bid to let the space as a café may have been more successful if the café had a separate entrance to the main health centre and was more outward-facing.*

modal expressions:

(18) *For example, the space behind the sculpture could be used for outdoor seating in the summer and passers-by would be able to see that there was a café available in the area.*

(19) *It would possibly be more appropriate to situate the health centre to face towards the south-west ...*

negated expressions or expressions noting absence:

(20) *There is no sign of art works in the corridors and stair cases creating a very monotone environment.*

(21) *... the lack of these views is a flaw in the design of the health centre.*

and finally conditional expressions noting things at could have been done differently or could be altered in the future.

(22) *If more green features such as trees and plants had been used in the court yards a better feeling of being in nature could have been encouraged in users of the building.*

(23) *However, if in the future the courtyards are made available to patients this may become a more serious problem.*

## 4. Annotating Design Evaluation Reports with ISO-Space

To better understand issues relating to the application of the ISO-Space annotation specification to building design evaluation reports, we began by selecting two of the four reports in our corpus to annotate. Using the ISO-Space annotation specification described in Pustejovsky et al. (2011a; 2011b), a human annotator added ISO-Space markup for locations, spatial entities and spatial signals to the reports, adhering to the guidelines as stricly as possible. In total, two reports were so annotated. Table 1 contains summary statistics of the annotated data. Following this we reviewed the annotations in the light of our intended application and made the following observations, some of which lead to proposals to extend or adapt ISO-Space for use in a broader range of applications.

### 4.1. Location vs. Spatial Entity

In ISO-Space a key distinction is that between *location* and *spatial entity*[7]. A location is characterised as "an inherently

[7] In SpatialML, what became the ISO-Space location element was a place element. In the latest unpublished version of the ISO-Space specification, version 1.4c, locations have been subdivided into two sorts, place and path, so "place" has reappeared at the preferred term for what Pustejovsky et al. (2011a), and we in the following, will refer to as location (Pustejovsky, personal communication, 2012).

| Feature | Count |
|---|---|
| Words | 13 052 |
| Spatial entities | 503 |
| Locations | 26 |
| Spatial signals | 84 |
| Events | 6 |
| Motions | 2 |

Table 1: Summary of ISO-Space element counts in annotated documents

grounded spatial entity", with exemplars being things like countries, mountains, cities and rivers (Pustejovsky et al., 2011a) – the sorts of named things one typically finds in geographical gazetteers. By contrast a spatial entity is "an entity that is not inherently a LOCATION, but one which is identified as participating in a spatial relation", examples being *car*, *building* or *John* or event-like things such as *traffic jam* or *hurricane*. Further to this, Pustejovsky et al. (2011a) say:

> Each SPATIAL_ENTITY inherently defines a location and can be the location for other spatial entities, as in *John is in the car*. This raises the issue of whether entities like *building* in *The statue is in the building* are annotated as locations or spatial entities. We resolve this by stipulating that these entities are never annotated as locations but always as spatial entities, even in a case like *the president is in the building*.

Following this instruction, in our first pass at annotating two of the Jordanthorpe design evaluation reports, we annotated all references to the building and to parts of it or things within it as SPATIAL_ENTITIES. This led to the relative proportions of these two types seen in Table 1.

However, on reflection we began to question whether this was the correct choice. If we ask what "inherently grounded" (the proposed defining characteristic of locations) means we see that there is no straightforward answer. If it means "has a fixed set of geospatial co-ordinates over an extended time period", then there is the difficulty of specifying precisely how long the extended time period should be. If it is too long then this definition fails to admit things that would seem to be locations, such as mountains and islands in Iceland that have formed recently (e.g. Surtsey Island, formed in 1963, or Eldfell, the mountain formed in 1973) and furthermore there will be difficulties with continental drift which, over an extended time period, leads to geospatial coordinates of landscape features, such as mountains and rivers – things which we might normally unquestioningly think of as locations – changing. On the other hand if a time period which is too short is chosen then various things, such as very old trees or ancient monuments (Stonehenge, Westminster Abbey), would seem to qualify. Furthermore extra-terrestrial bodies such the moon or indeed the rest of the Universe, which have a good claim to be considered locations, are also excluded. Another analysis might be "has an entry in a geographical gazetteer". This also seems unsatisfactory in that gazetteers may well be missing some entries, unnamed geographical features that are just like others which do have names and are in gazetteers get excluded, celestial locations are again excluded, and so on.

The task of providing a philosophically satisfactory account of the difference between locations and spatial entities is indeed a challenging one, and not one that we are going to attempt. However, we wish to advance a pragmatically motivated proposal that we believe usefully generalises the ISO-Space model. Rather than assume, as the current ISO-Space model appears to do, that a location is something that is fixed in space and across time, a more flexible approach would be to acknowledge that what constitutes a location will vary depending on the spatial and temporal scale adopted in a particular discourse. We believe that two two related, fundamentally sound intuitions about the distinction between locations and (other) spatial entities are as follows:

1. locations are (relatively) positionally stable entities in the spatial frame of reference for the discourse we are trying to analyze, whereas other spatial entities tend to move about within the frame of reference in a time scale during which the locations remain fixed;
2. the sort of things that appear in gazetteers are the names of locations (at the spatial and temporal scale for which the gazetteer is appropriate – presumably because it is the names of the things that are relatively positionally stable at the appropriate scale that find their way into gazetteers).

We also believe that in the context of semantic annotation there are two requirements on any distinction between tagged elements:

1. any distinction between classes of annotated textual elements should serve some purpose in some intended application of the annotation;
2. any distinction should be clear enough that annotators can easily and reliably recognize it.

One obvious purpose that locations serve in the sorts of applications used to motivate SpatialML and ISO-Space is that of being the entities linked to gazetteers and to geo-coordinates. I.e. it is locations that allow texts to be linked to other, graphical forms of representation via links to gazetteer database entries or via geo-coordinates. In our application scenario, the linking we are interested in is that from building elements to a 2D or 3D graphical model of the building. Thus, the pragmatic position we take is that we should allow building elements that can be mapped via the sort of "building gazetteer" mentioned above in Section 2. (rooms, stairways, named functional areas and so on) to be locations. This fits with our intuitions that locations be relatively stable and be the sorts of things whose names appear in gazetteers or maps of some terrain. By contrast, spatial entities are things that may move around in locations (such as furniture, art work and plants).

Thus, our first proposal to extend ISO-Space to support a broader range of applications is to clarify the distinction between locations and spatial entities, clarify the role that grounding plays in identifying locations and allow grounding not just at the scale of geographic features and geocoordinates, but at whatever scale is appropriate for linking the text to 2D or 3D co-ordinate or map/model based representation of the spatial world being described in the text and for which external models exist. Concrete suggestions on how to do this within the syntax of ISO-Space are discussed in the next section.

### 4.2. Multiple Scales and Frames of Reference

As discussed above in Section 3.1., a single building design evaluation report will frequently talk about the spatial properties of buildings at different scales. Thus, the site of the building within the broader urban or geographical setting will be discussed, as will, e.g., details of room positions within the building.

While descriptions at different scales may be specified within the same co-ordinate system, they need not be. Thus, a building might be located within a city using lat-long coordinates, but a BIM or CAD representation of the interior might take as an origin some arbitrary point within the building (e.g. lower left corner when viewed from the front). Building elements, such as room positions, for example, within the building will then be defined in the building co-ordinate system.

In this case, if the building co-ordinate system origin can be given a lat-long co-ordinate, perhaps via a nearby survey benchmark, then the building co-ordinate system can be embedded in the geo-co-ordinate system. In general, however, there is no reason why the multiple co-ordinate systems referenced in a document need to be such that one is embeddable in the other. The relation between them might be unspecified or the spaces they define may be non-intersecting or moving in relation to each other (imagine a story that alternates between describing activities on a planet's surface and on a space station orbiting the planet).

The term "frame of reference" is used in physics to describe a co-ordinate system which can be used to describe the position and motion of entities within it. Frames of reference can be embedded within each other or in motion in relation to each other. In order to deal with texts that involve multiple frames of reference, or even single frames of reference other than the base geo-co-ordinate frame of reference assumed currently in ISO-Space, we believe the ISO-Space model needs to be generalized to incorporate some such notion. Of course, "frame of reference" is already used in ISO-Space, and more broadly by cognitive linguists, to distinguish the types of orientation relation that are found in language systems – i.e. absolute, relative or intrinsic. We are not arguing to replace the linguistic usage with the physics one in ISO-Space, just to point out that this other sense is also relevant and should be incorporated into a framework for talking about the multiple levels of spatial description which occur in certain document types.

If we think of a gazetteer (or something like a map, floor plan or 3D model) and its associated co-ordinate system as providing a "frame of reference", then we need to be able to associate more than one frame of reference with a document. In fact, it is individual locations that are associated with frames of reference and therefore we need in principle to be able to associate a distinct frame of reference with every location mentioned in the text. In our view, therefore, the specification of the attributes associated with a location should be generalized to include a frame of reference attribute (in the physics sense). Of course the range of attribute values to be associated with certain attributes of location entity will depend on the frame of reference chosen. So, for example, the gazetteer reference will be to a gazetteer appropriate for frame of reference and the admissible location types will be frame of reference dependent (while possible types at the geo-level are, e.g. "continent", "body of water", and so on, at the building level appropriate types might be, e.g. "room", "corridor", etc.). These changes could be accommodated with relatively minor alterations to the existing syntax of ISO-Space.

### 4.3. Spatial Expressions in Intentional, Modal, Negated and Conditional Contexts

As noted in Section 3.6., there is a rich selection of cases where spatial expressions occur within intentional, modal, negated or conditional contexts. While these contexts are by no means predominant in the design evaluation reports, they occur sufficiently frequently that they cannot be ignored. In particular any algorithm processing spatial expressions in such contexts cannot assume either that the locations or spatial entities mentioned within them exist (though they may) or that the relations proposed between them actually hold. So, for example, in example (17), if *separate entrance to the main health centre* were tagged as a location, there would be no point trying to ground it in relation to the building gazetteer, because the entrance does not exist; in example (19) the health centre and the south-west are bona fide locations, but the orientation relation mentioned (*face towards*) does not hold between them. On the other hand such contexts may include spatial expressions that denote real locations and grounding them is important for our intended application of linking the reports to a 2D or 3D graphical representation to help readers better understand the text by "seeing" the context. For example, example the blinds mentioned in example (16) above are introduced two sentences earlier in the text by the negated intentional construct

(24) *There did not appear to be any blinds available to cover the high level windows and the double-height glazing at the end of the waiting area and surrounding the courtyards.*

but here the spatial entities (*the high level windows and the double-height glazing*) and locations (*waiting area* and *courtyards*) are real and should be grounded to allow an application to display a view of the relevant portion of the building.

At present there is no facility within ISO-Space to deal with these cases, though the problem has been noted in Pustejovsky et al. (2011a) as a topic for future work. Similar issues arose and have been addressed in the development of TimeML (Pustejovsky et al., 2003), one of the standards contributing to ISO-Space, for the related problems of annotating temporal expressions and events within negated, modal, conditional and intentional contexts. We do not attempt to review that work here, but believe that parts of the solution developed there can be re-used to address some of the problems highlighted here. In particular the SLINK tag which was used in TimeML to mark sub-ordinated contexts, i.e. modal, conditional and intentional contexts, could be used here as well. Tagging such sub-ordinated contexts at least serves to flag the fact that spatial expressions and relations within these contexts need to be treated specially, as they may not reflect what is the case. At this point we do not have an analysis that distinguishes those spatial expressions within sub-ordinating contexts that do genuinely refer to those that do not – this problem remains to be investigated.

Aside from sub-ordinated contexts, there are also straightforward cases of negation – see examples (20) and (21). These frequently reflect the non-existence of a spatial entity, e.g. *no blinds*. For such cases, a simple solution might be to add a POLARITY attribute to the *spatial_entity* tag, in the way that TimeML associates a POLARITY attribute with the EVENT tag (this can be one aspect of a more general similarity between spatial entities and events as things that occur in space and time respectively). Less clear is how to handle references to absence of functional spaces in cases like *there is no small waiting area for those who require privacy* or to abstract spatial entities like views (*lack of views*).

In sum we propose that the ISO-Space specification address sub-ordinating contexts containing spatial expressions by explicitly confirming the inclusion of the SLINK tag. Further we suggest that a POLARITY attribute be considered as a mechanism to address assertions of the non-existence of a spatial entity. More work remains to be done to analyze difficult cases of negation and sub-ordination.

### 4.4. Identity and Coreference

Given that the focus of design evaluation reports is frequently on spatial aspects of the buildings being evaluated, locations and spatial entities are frequently in grammatically focal positions in sentences and are referred to across multiple sentences. This introduces all the well-known problems of coreference in natural language texts, including anaphora, varying definite descriptions, etc. Linking these multiple references to the same entities is essential for understanding, for example, what part of a building an evaluative statement may refer to (for example *the centre* and *the building* in examples (10) and (13) respectively cannot be grounded with recognizing that they refer to the Jordanthorpe Health Centre.

At present there appears to be no way to link multiple references to the same location or spatial entity. The closest relation in the current ISO-Space specification is RCC8 EQ for "equal" (Randell et al., 1992). However, this relation is ambiguous as to whether the entities it co-ordinates are the same object or are separate entities have the same spatial bounds. An identity relation would support co-referential spatial descriptions, and disambiguate two mentions of the same object from two objects with the same bounds. A similar problem was encountered in TimeML where a distinction needed to be made between distinct but *simultaneous* events and multiple references to the same event. There it was solved by introducing an IDENTITY relation type in addition to an SIMULTANEOUS relation type. A similar solution could be adopted in ISO-Space by, e.g. adding an ID relation type, distinct from the EQ relation type already present in the RCC8 set, to the set of allowable relation types on the qualitative spatial link tag.

## 5. Related work

The closest prior work on spatial annotation at the level we have investigated in this paper is by Blaylock (2011) who explores the general problem of describing street-level objects and events. Prior work on automatic annotation of both entities and links has examined mostly geographical entities (Mani et al., 2008) and generic approaches to spatial relation labelling (Shen et al., 2009). Some previous work has been carried out on linking spatial descriptions to visual representations in Barker and Purves (2008), who address the problem of analyzing photo captions in order to geo-reference the image.

Extracting terminology related to buildings has been examined thoroughly (Meyer, 2001). Wonka et al. (2003) presents a formal building construction grammar based on English terminology. Recently there have been efforts to build ontological models of architectural and construction related concepts and terms (Eliseo et al., 2011; Bhatt et al., 2011).

## 6. Conclusion

In this paper we have introduced a novel application scenario for the annotation of spatial information in texts – the annotation of design evaluation reports for health care facilities. On the practical side this scenario is motivated by the desire to link texts to graphical representations such as maps or 3D models in order to improve their comprehensibility and to support novel access and summarization capabilities. However, aside from being a compelling application scenario, this scenario offers new challenges for standards for spatial annotation such as ISO-Space because the documents in the domain are so rich in spatial language and because the scenario requires the application of the standard at a scale not yet investigated. We illustrated this challenge by cataloguing some of the wide range of spatial language found in design evaluation reports. We went on to describe preliminary work on annotating several design evaluation reports using ISO-Space. This effort exposed some fundamental issues that arise when applying the ISO-Space specification to documents discussing spatial locations, entities and relations at the scale of buildings and in the context of an application which requires grounding this information in an externally supplied model. From our analysis of these issues we proposed four extensions to the current ISO-Space specification:

1. a more nuanced description of the distinction between locations and spatial entities that will allow locations at other than the geo-centric scale implicit in the description of the standard so far and will enable clear-

cut decisions to be made by annotators, perhaps operationalising the distinction in terms of what can be grounded by reference to a specific external resource relating named entities to a co-ordinate system;

2. an explicit encoding of something like a frame of reference attribute that will support interpretation documents that contain spatial descriptions in multiple frames of reference – other attributes of locations, such as type and gazetteer reference, would then need to be interpreted in relation to the specified frame of reference, and their admissible values would depend upon that frame of reference;

3. an explicit acknowledgement that something like the TimeML SLINK should be used to identify spatial expressions that occur on sub-ordinated contexts, such as modal, intentional and conditional contexts and that something like the TimeML POLARITY attribute should be added to spatial entities so that assertions that deny their existence can be properly encoded;

4. the addition of something like an ID relation to the set of qualitative spatial link types, distinct from the EQ relation, in order to distinguish multiple references to the same spatial entity or location from references to multiple spatial entities or locations occupying the same place.

Turning to the future, our plan is to annotate fully our small corpus of design evaluation reports with an extended version of ISO-Space that takes into account the proposals above. There are no doubt additional challenges to be addressed in applying ISO-Space, as we move to add spatial links and to deal with the sorts of examples discussed in Section 3.2. on multiple perspectives, such as expressions referring to functionally specified areas (e.g. a *staff-only zone*). Following this we will begin to develop tools to support automated annotation and grounding and then to integrate the language processing components with 3D graphical representations in order to address the complete application scenario. Finally, moving beyond design evaluation reports there is a huge range of other similar applications relating textual documents to designed objects.

**Acknowledgments**

The authors thank James Pustejovsky for helpful discussions regarding the ISO-Space standard. The authors would also like to acknowledge support of the University of Sheffield in the form of a Research Stimulation Fund grant. Finally, the third author acknowledges the support of the UK Engineering and Physical Science Research Council in the form of a doctoral training grant.

## 7. References

E.J. Barker and R.S. Purves. 2008. A caption annotation scheme for georeferencing images. In *Proceedings of the International Workshop on Geographic Information Retrieval*, pages 45–46. ACM.

M. Bhatt, J. Hois, and O. Kutz. 2011. Ontological Modelling of Form and Function for Architectural Design. *Applied Ontology*, 1:32.

N. Blaylock. 2011. Semantic Annotation of Street-level Geospatial Entities. In *Proceedings of the IEEE International Conference on Semantic Computing*, pages 444–448. IEEE.

M.A. Eliseo, J.M.P. de Oliveira, and S.R.M. Pellegrin. 2011. Domain Ontology with Temporal Descriptions for Architectural Buildings as a Support for Learning History of Architecture. *IEEE Multidisciplinary Engineering Education Magazine*, 6:24–31.

I. Mani, J. Hitzeman, and C. Clark. 2008. Annotating natural language geographic references. In *Proceedings of the LREC Workshop on Methodologies and Resources for Processing Spatial Language*.

I. Mani, C. Doran, D. Harris, J. Hitzeman, R. Quimby, J. Richer, B. Wellner, S. Mardis, and S. Clancy. 2010. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44:263–280. 10.1007/s10579-010-9121-0.

I. Meyer. 2001. Extracting knowledge-rich contexts for terminography. *Recent advances in computational terminology*, pages 279–302.

W.F.E. Preiser, H.Z. Rabinowitz, E.T. White, et al. 1988. *Post-occupancy evaluation*. Van Nostrand Reinhold New York.

J. Pustejovsky, J. Castaño, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, and G. Katz. 2003. TimeML: Robust specification of event and temporal expressions in text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg, January.

J. Pustejovsky, J.L. Moszkowicz, and M. Verhagen. 2011a. ISO-Space: The annotation of spatial information in language. In *Proceedings of the Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, pages 1–9.

J. Pustejovsky, J.L. Moszkowicz, and M. Verhagen. 2011b. Using ISO-Space for Annotating Spatial Information. In *Proceedings of the International Conference on Spatial Information Theory*.

D.A. Randell, Z. Cui, and A.G. Cohn. 1992. A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, pages 165–176.

Q. Shen, X. Zhang, and W. Jiang. 2009. Annotation of spatial relations in natural language. In *Proceedings of the International Conference on Environmental Science and Information Application Technology*, pages 418–421. IEEE.

UK Department of Health. 2008a. A Staff and Patient Environment Calibration Tool (ASPECT). Technical report, Department of Health, UK.

UK Department of Health. 2008b. Achieving Excellence Design Evaluation Toolkit (AEDET Evolution). Technical report, Department of Health, UK.

P. Wonka, M. Wimmer, F. Sillion, and W. Ribarsky. 2003. Instant architecture. In *ACM SIGGRAPH 2003 Papers*, pages 669–677. ACM.

# Treebanks of Logical Forms: they are Useful Only if Consistent

**Rodolfo Delmonte and Agata Rotondi**

Department of Language Studies, Università Ca' Foscari - Ca' Bembo, 30123 – Venezia (It)
E-mail: delmont@unive.it Website: project.cgm.unive.it

**Abstract**

Logical Forms are an exceptionally important linguistic representation for highly demanding semantically related tasks like Question/ Answering and Text Understanding, but their automatic production at runtime is higly error-prone. The use of a tool like XWNet and other similar resources would be beneficial for all the NLP community, but not only. The problem is: Logical Forms are useful as long as they are consistent, otherwise they would be useless if not harmful. Like any other resource that aims at providing a meaning representation, LFs require a big effort in manual checking order to reduce the number of errors to the minimum acceptable – less than 1% - from any digital resource. As will be shown in detail in the paper, the available resources – XWNet, WN30-lfs, ILF - suffer from lack of a careful manual checking phase, and the number of errors is too high to make the resource usable as is. We classified mistakes by their syntactic or semantic type in order to facilitate a revision of the resource that we intend to do using regular expressions. We also commented extensively on semantic issues and on the best way to represent them in Logical Forms.

## 1. Introduction

In a number of recent papers, the need for a sizable (at least same size of WordNet) and publicly available corpus with Logical Form representation has increased: as a result more and more papers are concerned with the generation of a logical form or a semantic representation that is close to it. The fact is that there is already a number of such resources available, XWN (Moldovan and Rus, 2001), and ILF (Agerri and Peñas, 2010), hence (AP), both derived from WordNet glosses: so, why not using them. In fact in their paper, after reviewing previous work - including XWN and WN30-lfs (by Clark et al., 2008) generated by USC/ISI, California in 2006 - AP come to the conclusion that "... there is still some need for providing lexical and/or knowledge resources suitable for computational semantics tasks that required formalized knowledge." (ibid.29) The problem seems to be the presence of some obscurity in the way in which the glosses have been transformed - WN30-lfs is commented as containing "... free variables and/or predicates without any relation with any other predicates in the definition"(ibid.29) and the same problem is also present in XWN2 (ibid.,28). Here in addition, the output is cluttered with elements of the gloss which do not contribute to the definition strictly speaking, that is examples coming with the gloss. In fact also Clark et al. complain about the lack of consistency of XWN but no details are given.

Of course not all published comments on XWN speak negatively - without any detailed analysis, in fact - of XWN: on the contrary all published work by the authors of XWN speaks in favour of it. There are many papers published by the authors, V.Rus, D.Moldovan, S.Harabagiu et al., R.Mihalcea et al. – see the References -, who describe their work positively, if not highly positively, and comment on its usefulness for various semantically heavy tasks like Question Answering and RTE. In particular, Rus indicated an experiment with evaluation, where the accuracy for glosses conversion into Logical Forms is reported at 89.46%(Rus V., 2001), but on a selection of 1000 WN glosses only. The conclusion would be an error rate slightly over 10%, which is an important quantity of data but still perhaps bearable. In fact, we found over 30% error rate, and this is why – in our opinion - the XWN is badly flawed and cannot be used for the purpose it was made.

In the following sections we will go through the typical mistakes present in the corpus and comment on them. We don't want to imply that work carried out is useless, but since it can improved we intend to correct it in the future, and provinding classes of mistakes seems to best way to help doing that. A lot of difficult problems have been solved in XWN that deserve the resource to be saved and improved upon. Producing such a resource from scratch is outside the scope of current NLP technology, and this is attested by the various attempts at achieving such a goal (see also Ovchinnikova et al., 2011). However, there are also other attempts at producing Logical Forms directly from Penn Treebank style syntactic representations, like for instance, the LFToolkit by Nishit Rashod and Jerry Hobbs at their website, and the experiment reported by Alshawi et al. that we comment on here below.

In Alshawi et al. (2011) an experiment is reported to derive sentence-semantics pairs for training and testing from the Penn Treebank. In order to do that they program the Stanford treebank toolkit to produce what they call NLF expressions, that is Natural Logical Form, which are intentionally not intended as fully resolved logical forms. These are meant to be closer to natural logic than QLF Quasi Logical Forms, in order to use them to make some Natural Logic inference. And as the authors themselves comment, QLFs are being used widely to refer to any logic-like semantic representation without explicit quantifier scope, i.e. unscoped logical forms(ibid.17). In the same paper the authors specifically comment on the need to use an unknown/unspecified Null operator, %, for all those linguistic constructs which are beyond the coverage of their semantic model. This applies to a great number of constructions that are present in the PTB and they give slightly different results in accuracy, both around 86%, however. Here again, we have to note that the usefulness of such logic-like representation is very low due to incompleteness of its results.

The Null operator is also present in PTB for all those linguistic constructions that have been regarded too difficult to take decisions upon by annotators and include all adjunct infinitivals and gerundives for a total amount of some 12,000 non coindexed null elements. This problem has also prevented other attempts at producing a semantically viable corpus of logical forms directly from a mapping of PTB, by a number of other researchers working in the LFG framework, (Guo et al.,2007) and in HPSG and CCG frameworks, but also Dependency Grammar as reported in (Nivre and Nilsson, 2005).

All these methods go beyond the encoding of surface context-free phrase structure trees, to incorporate non-local dependencies. This option requires recovering empty nodes and identifying their antecedents, be they traces or long distance dependencies. But since PTB annotators themselves intentionally refused to coindex all those cases that caused some difficulty in the decision process, all work carried out on this resource is flawed, semantically speaking, from the start. We must however, admit to the fact that WN glosses are much simpler sentences in comparison to PTB sentences, which even if taken with a word limit under 40 are still too complex and not comparable to definitions.

## 2. Common Mistakes and Their Classification

Logical Forms in XWN are graded in three quality levels: normal, silver and gold; the same applies to tagging and phrase structure constituency. "Normal" quality, which applies to the majority of the glosses, is used to indicate that there is no agreement between the two parsers that have been used to parse the input definition, and that there has been no manual checking of the output. "Gold" quality means manual checking has been performed, and "silver" quality indicates that there has been no manual checking but the two parsers agree in their representation. The importance given to the agreement between the two constituency parsers, is due to the fact that LFs are a mapping on syntactic constituency representation.

LF from glosses is represented in different manner according to lexical category, adjective, verb, noun and adverb: each one is associated to a predicate but with some differences. We list here below examples for each category:

A. Nouns.
For each synset the argument 'x1' is assigned to the first word that it represents. In the gloss, the 'x1' variable is referred to the same entity of the first word in the synset, as in:

*plant:NN(x1) -> living:JJ(x1) organism:NN(x1) lack:VB(e1, x1, x2) power:NN(x2) of:IN(x2, x3) locomotion:NN(x3)*

where *plant(x1)* and *living(x1) organism(x1)* refer to the same entity. One of the important efforts that characterizes positively XWN is the treatment of nominal compound, which has been done following Hobbs' suggestion in TACITUS to introduce the predicate NN in LF. Predicates may have a variable number of arguments and only the first is associated to the aggregation or compound of all the composing arguments, as in
*jam_session:NN(x1) -> impromptu:JJ(x1) nn(x1, x2, x3) jazz:NN(x2) concert:NN(x3)*

B. Verbs.
For each synset, the variable 'e1' is associated to the first term that represents it, to indicate the eventuality of the action/state/event of the verb meaning; the subject is associated invariably to 'x1' and the object to 'x2'. The second argument may be fictitious in case of intransitive verbs.

*recognize:VB(e1, x1, x2) -> show:VB(e1, x1, x5) approval:NN(x3) or:CC(x5, x3, x4) appreciation:NN(x4) of:IN(x5, x2)*

In this case all variables are bound to some argument position and are associated to some linguistic element. In the following example, an intransitive verb, we see on the contrary that there are two fictitious objects:

*tremble:VB(e1, x1, x2) -> move:VB(e1, x1, x4) with:IN(e1, x3) tremor:NN(x3)*

In the case of ditransitive verbs, the LF representation of the event is verb(e1,x1,x2,x3), as in, *professor gives students the grades: professor(x1 ) give( e1, x1, x2, x3 ) grade(x2) student (x3),* or in the definition of the verb GIVE:

*give:VB(e1, x1, x2, x3) -> allow:VB(e1, x1, x3) to:IN(e1, e4) have:VB(e2, x3, x2) or:CC(e4, e2, e3) take:VB(e3, x3, x2)*

C. Adjectives.
For each synset, argument 'x1' is associated to the first word that represents it, then in the second part of the gloss, argument 'x1' refers to the same entity described by the first word in the synset, as in,

*ascetic:JJ(x1) -> practice:VB(e1, x1, x2) great:JJ(x2) self-denial:NN(x2)*

D. Adverbs.

For each synset, argument 'e1' is assigned to the first term that represents it, then in the second part of the gloss, argument 'e1' refers to the same action, as in

*grossly:RB(e1) -> in:IN(e1, x1) gross:JJ(x1) manner:NN(x1) largely:RB(e1) -> mainly:RB(e1) chiefly:RB(e1)*

Other categories are treated as follows: prepositions are treated as predicates with two arguments, the first being the head noun that is modified by the prepositional phrase, and the second being the modified head noun; possessive pronouns introduce a relation between the governing head and the referent of the possessive pronoun, the predicate POS is used to represent this relation. What LFs do not contain are: verbal tense and mood (which can be regarded less relevant in definitions), negation, quantifiers (they are treated as adjectives or pronouns) and modal operators, comparative operators, plural, gender, illocutionary force and speech acts. Some of these semantic markers are only present, however, in few cases, as for instance in (A. Ramsay and D. Field, 2008).

We report here below common mistakes we found in the LF representation of XWN. This work has been carried out trying to group the most common mistakes into classes, be they related to tagging, to syntactic structure, to lexical types or semantic types. Of course for lack of space, we will not be able to discuss mor than one example per mistake. The first type of mistakes regards the disappearance of CONJUNCTIONS in coordinate structures and the consequent lack of binding of logical variables: here below we report the gloss focussing on the important portion of it and disregarding additional information.

Case 1: CONJUNCTIONS

Here, the missing conjunction is OR, and the unbound variable is "x5", also note that the coordinating conjunction AND is assigned variables which do not have any correspondence in the representation.

<gloss pos="NOUN" synsetID="07164600">
<synonymSet>seedcake, seed_cake</synonymSet>
<text> a sweet cake flavored with sesame or caraway seeds and lemon
</text>
<lft quality="NORMAL">
seedcake:NN(x1) -> sweet:JJ(x1) cake:NN(x1) flavor:VB(e1, x7, x1) with:IN(e1, x6) sesame:NN(x2) caraway:JJ(x5) seed:NN(x3) and:CC(x30, x31, x32) lemon:NN(x4) </lft>
</gloss>

Case 2: PHRASAL VERBS AND PREPOSITIONAL VERBS

Most frequent prepositions appearing in the database are: on, in, to, by, for, with, at, of, from, as. Some of them have an anomalous behaviour in the LF in case they appear at the end of the gloss: they are sometimes erased, and this does not depend on the quality of the LF because this happens in all three types, silver, gold or normal.

<gloss pos="VERB" synsetID="00042006">
<synonymSet>perfume, scent</synonymSet>
<text> apply perfume to; "She perfumes herself every day"
</text>
<lft quality="GOLD">
perfume:VB(e1, x1, x2) -> apply:VB(e1, x1, x3) perfume:NN(x3)
</lft> </gloss>

Here, on the contrary it is preserved,

<gloss pos="VERB" synsetID="00040699">
<synonymSet>powder</synonymSet>
<text> apply powder to; "She powdered her nose"; "The King wears a powdered wig"
</text>
<lft quality="GOLD">
powder:VB(e1, x1, x2) -> apply:VB(e1, x1, x3) powder:NN(x3) to:IN(e1, x2) </lft>
</gloss>

As to phrasal verbs the treatment is not homogeneous and sometimes the verb particle may simply be erased. It can appear attached to the verb as in (work_out:VB), or in a separate entry (set:VB(e1,x3,x1) up:IN(e1,x2)), or simply disappear,

<gloss pos="NOUN" synsetID="07918617">
<synonymSet>secondary</synonymSet>
<text>the defensive football players who line up behind the linemen
</text>
<lft quality="NORMAL">
secondary:JJ(x4) -> defensive:JJ(x1) football:NN(x1) player:NN(x1) line:VB(e1, x1, x26) behind:IN(e1, x2) linemen:NN(x2) </lft>
</gloss>

Case 3: NOMINAL COMPOUNDS

As said above, nominal compounds are mapped into LF by means of the predicate *nn*. There is a great number of compunds which still have to be identified and mapped into LF, here however we refer to the case of a given compound which is identified but then it is mapped differently in different contexts. One such cases is constituted by "World War":

<lft quality="NORMAL">
bataan:NN(x1) -> peninsula:NN(x2) and:CC(x1, x2, x3) island:NN(x3) in:IN(x1, x4) philippines:NN(x4) japanese:JJ(x5)

force:NN(x5) besiege:VB(e1, x9, x5) american:NN(x6) force:NN(x7) in:IN(x6, x8) world_war_ii:NN(x8) </lft>

<lft quality="NORMAL">
wac:NN(x1) -> member:NN(x1) of:IN(x1, x2) women's:NN(x2) army:NN(x3) corp:NN(x4) be:VB(e1, x2, e2) organize:VB (e2, x9, x2) during:IN(e2, x5) world:NN(x5) war:NN(x6) ii:NN(x7) but:CC(e4, e0, e3) be:VB(e3, x1, x8) no:RB(e3) longer:RB(e3) separate:JJ(x8) branch:NN(x8) </lft>

<lft quality="SILVER">
battle_of_the_ardennes_bulge:NN(x1) -> battle:NN(x1) during:IN(x1, x2) world:NN(x2) war:JJ(x2) ii:NN(x3)
</lft>
<lft quality="NORMAL">
snafu:NN(x1) -> acronym:NN(x1) often:RB(e0) use:VB(e1, x2, x1) by:IN(e1, x2) soldier:NN(x2) in:IN(e1, x3) world:NN (x3) war:NN(x4) ii:JJ(x3) situation:NN(x5) normal:JJ(x6) all:JJ(x6) fucked:NN(x6) up:IN(e1, x6) </lft>

As can be noticed, the component words of "World War II" are analysed alternatively as separate Nouns, Nouns and Adjective, or as a single Noun.

## Case 4: TAGGING ERRORS

The most frequent mistake in each of the four separate files is certainly the wrong POS assigned by the tagger. However in some cases the syntactic tree contains the right category while the LF has a corresponding wrong one.

<gloss pos="NOUN" synsetID="10317346">
<synonymSet>Hawking, Stephen_Hawking, Stephen_William_Hawking</synonymSet>
<text> English theoretical physicist (born in 1942)
</text>
<lft quality="NORMAL">
hawk:VB(e1, x3) -> english:NN(x1) theoretical:JJ(x1) physicist:NN(x2)</lft>
</gloss>

<gloss pos="ADV" synsetID="00288722">
<synonymSet>clear, all_the_way</synonymSet>
<text> completely; "read ..."
</text>
<lft quality="GOLD">
clear:JJ(e1) -> completely:RB(e1)</lft>
</gloss>

Here the surname "Hawking" is turned into the verb "HAWK", then the ADVerbial "completely" is associated to an adjective JJ CLEAR. As will be commented below, there are many problems in the encoding of participles, as shown here again,

<gloss pos="NOUN" synsetID="00209984">
<synonymSet>chance-medley</synonymSet>
<text> unpremeditated killing of a human being in self defense
</text>
<lft quality="NORMAL">
chance-medley:NN(x1) -> unpremeditated:VB(e1, x5, x1) killing:NN(x1) of:IN(x1, x2) human:NN(x2) in:IN(x2, x3) self:NN(x3) defense:NN(x4) </lft>
</gloss>

and here,

<gloss pos="NOUN" synsetID="09420441">
<synonymSet>esthetician, aesthetician</synonymSet>
<text> a worker skilled in giving beauty treatments (manicures and facials etc.)
</text>
<lft quality="NORMAL">
esthetician:NN(x1) -> worker:NN(x1) skilled:VB(e1, x4, x1) in:IN(e1, e2) give:VB(e2, x1, x2) beauty:NN(x2) treatment:NN(x3) </lft>
</gloss>

Gerundives or present participles, when appearing at the beginning of a definition, are mapped onto the verb base form preceded by "act of", as in

advancing toward a goal -> act:NN(x1) of:IN(x1, e1) advance:VB(e1, x2, x26) toward:IN(e1, x2) goal:NN(x2).

However, this should not happen when the –ing form is used as a nominalized verb as in

<gloss pos="NOUN" synsetID="05877558">
<synonymSet>notepaper</synonymSet>
<text> writing paper intended for writing short notes or letters
</text>
<lft quality="SILVER">
notepaper:NN(x1) -> act:NN(x1) of:IN(x1, e1) write:VB(e1, x2, x2) paper:NN(x2) intend:VB(e2, x6, x2) for:IN(e2, e3)
write:VB(e3, x2, x5) short:JJ(x5) note:NN(x3) or:CC(x5, x3, x4) letters:NN(x4) </lft>
</gloss>

Genitive marking is interpreted in many different ways, as a Noun, Adjective or even Verb, in the Noun file, as shown here,

<gloss pos="NOUN" synsetID="00157666">
<synonymSet>capture</synonymSet>
<text>vthe removal of an opponent's piece from the chess board
</text>
<lft quality="NORMAL">
capture:NN(x1) -> removal:NN(x1) of:IN(x1, x2) opponent:NN(x2) 's:VB(e1, x2, x3) piece:NN(x3) from:IN(x3, x4) chess:NN(x4) board:NN(x5) </lft>
</gloss>

## Case 5: FREE VARIABLES

Indexed variables are fundamental element of the LF and are used to indicate relations intervening between event and arguments or modifiers. In some cases, fictitious arguments can appear with free variables at the event level, however when the argument is actually present - in particular, in intransitive or passivized structures -, it should be coindexed with the event. Very often this does not happen,

<lft quality="GOLD">
hibernate:VB(e1, x1, x2) -> sleep:VB(e1, x1, x9) during:IN(e1, x3) winter:NN(x3) </lft>
</gloss>

<text> a man of such superior qualities that he seems like a deity to other people; "he was a god among men"
</text>
<lft quality="NORMAL">
god:NN(x1) -> man:NN(x1) of:IN(x1, x2) such:JJ(x2) superior:JJ(x2) quality:NN(x2) that:IN(e1, x5) seem:VB(e1, x2, x26) like:IN(e1, x3) deity:NN(x3) to:IN(x3, x4) other:JJ(x4) people:NN(x4) </lft>

<text> a commissioned officer in the United States Army or Air Force or Marines holding a rank above major and below colonel </text>
<lft quality="NORMAL">
lieutenant_colonel:NN(x1) -> commission:VB(e1, x11, x1) officer:NN(x2) in:IN(e1, x9) united_states_army:NN(x3) or:CC (x9, x3, x1, x4) air:NN(x1) force:NN(x4) or:CC(e3, e1) marine:NN(x5) hold:VB(e2, x5, x6) rank:NN(x6) above:IN(e2, x10) major:JJ(x8) below:IN(x20, x21) colonel:NN(x7) </lft>

Case 6. : NEGATION
There are lots of negations in WN glosses – 3107 cases of NOT overall - and as we will see, a number of them are wrongly scoped, some 20%. In particular, negation is distributed as follows in the four files: 2024 in Adjectives; 947 in Nouns; 79 in Adverbs; 57 in Verbs. If we add the other negation markers (NO, NONE, NOTHING, NEVER, NOR) adding up to 676 occurrences, we come up with some 3783 cases.
Negation can receive different scope according to its semantic role: it can negate the main verb or modifiers of the verb like adverbials – and in this case it will receive wide scope over the proposition, verb and arguments - or it can negate some specific argument or adjunct and in this case it will receive narrow scope. The majority of the cases of narrow scope negation is present in the Adjectival file: there are 901 cases of wide scope – that is the gloss is expressed by a full proposition with a verb and some argument; then there 1095 cases of narrow scope which is all corretly marked, as shown here below:

absolute:JJ(x1) -> not:RB(x1) limited:JJ(x1) by:IN(x1, x2) law:NN(x2)

Besides, consider the case of "alien" with the meaning of "foreign", where the negation has wide scope of the coordination of two verbs,

alien:JJ(x1) -> not:RB(e3) contain:VB(e1, x7, x1) in:IN(e1, x5) or:CC(e3, e1, e2) derive:VB(e2, x1) from:IN(e2, x2) essential:JJ(x2) nature:NN(x2) of:IN(x2, x3) something:NN(x3)

This is done extensively over all the dataset. Most errors derive from the wrong mapping of syntactic information in most of the case in which the negation is attached to an auxiliary verb, HAVE, BE, DO. In all these cases, the mapping wrongly produces two event variables, one for the auxiliary and another for the main verb, and the scope of negation is assigned narrow scope over the event variable of the auxiliary, as shown here below,

absentee_rate:NN(x1) -> percentage:NN(x1) of:IN(x1, x2) worker:NN(x2) do:VB(e1, x2, e2) not:RB(e1) report:VB(e2, x2, x26) to:IN(e2, e3) work:VB(e3, x2, x26)

However, in some cases the scope is marked correctly on the main verb as in,

lowbrow:JJ(x1) -> characteristic:JJ(x2) of:IN(x1, x2) person:NN(x2) be:VB(e1, x2) not:RB(x5) cultivated:JJ(x5) or:CC(e4, e1, e2) do:VB(e2, x2, e3) not:RB(e3) have:VB(e3, x2, x3) intellectual:JJ(x3) taste:NN(x3)

## 3 Some general considerations on XWN

Some general considerations over the whole dataset come from considering the amount of GOLD data with respect to NORMAL or SILVER, as shown in Table 1.

| **Types** | **Adverb.** | **Adjectiv.** | **Verbs** | **Nouns** |
|---|---|---|---|---|
| Gold | 3994 | 16059 | 14441 | 32844 |
| Silver | 0 | 4321 | 0 | 7228 |
| Normal | 0 | 0 | 0 | 54796 |
| Total | 3994 | 20380 | 14441 | 94868 |

Table 1.: Number of Gold/Silver/Normal LF entries in XWN

As can be easily gathered, the number of errors will vary substantially from one file to the other depending strictly on the number of GOLD LF entries, and will be proportional to the overall size of the file in terms of total number of entries. The file in which most errors are found is the one of NOUNS, which is not only the only file to contain Normal entries, but also in a quantity which is much higher than the GOLD ones, almost the double. Another important factor that may be considered as possible cause of errors in the NOUN file is the length of the gloss in number of words, which is more extended in syntactic terms than in the other files.

As a final remark, we extracted all the records containing just the LF from every single file, we then sorted them and checked for their consistency: this was done in order to verify that no two Logical Form are identical to each other. Whenever this happens, the meaning associated to one synset would be interchangeable with the meaning associated to another synset, which is clearly a sign of inconsistency. We found the following situation,

- over 94868 entries for Nouns 43 are duplicate LFs
- over 20380 entries for Adjective, 47 are duplicate LFs
- over 3994 entries for Adverbs, 12 are duplicate LFs
- over 14441 entries for Verbs, 29 are duplicatre LFs

Here below we report some examples of duplicate, or sometimes triple LF representations taken from the Noun file:

alaska_peninsula:NN(x1) -> peninsula:NN(x1) in:IN(x1, x2) southwestern:JJ(x2) alaska:NN(x2)

alpaca:NN(x1) -> wool:NN(x1) of:IN(x1, x2) alpaca:NN(x2)

anagoge:NN(x1) -> mystical:JJ(x1) allegorical:JJ(x1) interpretation:NN(x1)

approbation:NN(x1) -> official:JJ(x1) approval:NN(x1)

bailey:NN(x1) -> outer:JJ(x1) courtyard:NN(x1) of:IN(x1, x2) castle:NN(x2)

Bernoulli:NN(x1) -> swiss:JJ(x1) mathematician:NN(x1)

blood_count:NN(x1) -> number:NN(x1) of:IN(x1, x2) red:JJ(x2) white:JJ(x2) corpuscle:NN(x2) in:IN(x2, x3) blood:NN(x3) sample:NN(x4)

card_catalog:NN(x1) -> enumeration:NN(x1) of:IN(x1, x2) resource:NN(x2) of:IN(x2, x3) library:NN(x3)

cassava:NN(x1) -> source:NN(x1) of:IN(x1, x2) tapioca:NN(x2)

catapult:NN(x1) -> use:VB(e1, x2, x1) to:IN(e1, e2) propel:VB(e2, x1, x1) small:JJ(x1) stone:NN(x1)

clash:NN(x1) -> state:NN(x1) of:IN(x1, x2) conflict:NN(x2) between:IN(x2, x3) person:NN(x3)

## 4. Intermediate Logical Forms

In their paper (Agirre & Peñas, 2010) the authors present an automatic system that produces LFs from WordNet glosses using Stanford Parser and then mapping the output with typed dependencies into what they call ILFs. The important contribution of these two authors is the preprocessing phase of the glosses in order to make them concise and homogeneous as much as possible. To this aim, the authors eliminate all content between parenthesis; they also eliminate all that comes after a semicolon. Then they treat the three main categories as follows: they add a period at the end of the gloss for all categories; nouns and adverbs have the first word capitalized; in the case of adjectives, they add the word "Something" at the beginning of the gloss; and with verbs, they add the particle "To" at the beginning.

If we compare the result obtained in ILF with the LFs of XWN we notice that the mistakes that we found and commented above still occur but with a much lower frequency. The most important mistake we noticed in XWN here it is totally absent: there are no unbound variables in LF, all the variables are bound regularly. In addition to XWN ILF contains all article, conjunctions and prepositions.

Overall, we have noticed a remarkable improvement in the LF representation but as the authors themselves comment, the resource needs improvement. In particular there is no word sense assigned uniquely to each gloss as happened in XWN. This could be easily amended given the availability of a newly released version of the glosses with sense disambiguation at WordNet website. The resource still needs some comprehensive evaluation and, as the authors themselves indicate, this will be done when version 1.0 will be available.

The first observation to be made is that the reduction and also the fact that the authors managed to focus on the definition and eliminated most if not all of the remaining additional unessential parts, is certainly to be judged positively. However, as we show below, the resulting Logical Form has on the contrary become less readable if not unreadable and difficult to use, in one word it has lost perspicuity. Consider one example:

<sense offset="301890382" pos="s"
synset_name="bigheaded.s.01"> <gloss>
<text>Something overly conceited or arrogant.</text> <parse
parser="Stanford parser 1.6.1">
…….
<ilf version="0.2">[rel(1,3,2,'advmod',G1_3,G1_2),
rel(1,1,3,'amod',G1_1,G1_3), rel(1,1,5,'amod',G1_1,G1_5),
rel(1,3,5,'conj_or',G1_3,G1_5), e(1,2,G1_2),
w(1,2,'overly','r','rb'), e(1,3,G1_3), w(1,3,'conceited','a','jj'),
syn(1,3,301891773), e(1,1,G1_1), w(1,1,'something','n','nn'),
e(1,5,G1_5), w(1,5,'arrogant','a','jj'),
syn(1,5,301889819)]</ilf>
<pretty-ilf>something(x1) amod(x1,x3) amod(x1,x5) overly(x2)
conceited(x3) advmod(x3,x2) conj_or(x3,x5) arrogant(x5)
</pretty-ilf>

The authors have cluttered the LF with all details derived from the dependency graph produced by Stanford's parser, including tags associated to words, dependency types, which are rendered as rel(ations) on the arc linking two words. Then words are eventually associated to their dependency indices but in addition, they have double tags, the ones produced by their tagger and the ones coming from Stanford's parser. Eventually they provide a "pretty print" version of ILF – for "easier readability" as they say (ibid.,33) – where a straightforward version appears with just words and variables. Here they use dependency types as prefixes: the net result is that the LF is populated by eight expressions just like what the original complete gloss would require. Here below we report the complete version of the LF produced in XWN, where the adjectives are simply treated as modifiers of the same head and the conjunction is erased,

<entry word="bigheaded#a#1" status="partial">
<gloss>used colloquially of one who is overly conceited or arrogant</gloss>
bigheaded:JJ(x1) -> use:VB(e1, x6, x1) colloquially:RB(e2) of:IN(e1, e2) one:JJ(x3) be:VB(e2, x1) overly:RB(x4) conceited:JJ(x4) arrogant:JJ(x4)

As can be noticed, ILF has reduced the linguistic content of the gloss but in so doing it has deleted important information regarding the register of usage of the main entry word marked as "colloquial". Also, the introduction of dependency types has made the overall LF representation less perspicuous and certainly difficult to use in practical applications.
Mistakes we found are as follows:
- not all contents within parenthesis have been eliminated:

<text>Pure ethyl alcohol (containing no more than 1% water).</text>
<pretty-ilf>pure(x1) ethyl(x2) alcohol(x3) amod(x3,x1) nn(x3,x2) ((x4) nsubj(x4,x3) dep(x4,x5) contain(x5) advmod(x5,x6) dobj(x5,x12) no(x6) dep(x6,x10) more(x7) than(x8) advmod(x8,x7) 1(x9) quantmod(x9,x8) %(x10) num(x10,x9) water(x11) )(x12) nn(x12,x11)</pretty-ilf>

The use of the predicate "nn" for compound nouns has been improved and we checked that also for the compound commented above, "World War II", which is mapped correctly; this notwithstanding there are many unneeded uses of the predicate "nn" as for instance in "coarse tobacco",

<text>A strong coarse tobacco that has been shredded.</text>
<word ind="1" pos="DT">a</word>
<word ind="2" pos="JJ">strong</word>
<word ind="3" pos="NN">coarse</word>
<word ind="4" pos="NN">tobacco</word>
<pretty-ilf>a(x1) strong(x2) coarse(x3) tobacco(x4) det(x4,x1) amod(x4,x2) nn(x4,x3) rcmod(x4,x8) that(x5) have(x6) be(x7) shred(x8) nsubjpass(x8,x4) rel(x8,x5) aux(x8,x6) auxpass(x8,x7)</pretty-ilf>

where we assume that there might have been a tagging error. More tagging errors occur with colour nouns and past participles. Other mistakes come from wrong cases of pp_attachment as for instance in the following entry,

<text>The nonrandom movement of an atom or radical from one place to another within a molecule.</text>
<pretty-ilf>the(x1) nonrandom(x2) movement(x3) det(x3,x1) amod(x3,x2) prep_of(x3,x6) prep_of(x3,x11) prep_to(x3,x13) a(x5) atom(x6) det(x6,x5) conj_or(x6,x11) radical(x8) prep_from(x8,x10) one(x10) place(x11) amod(x11,x8) another(x13) prep_within(x13,x16) a(x15) molecule(x16) det(x16,x15)</pretty-ilf>

Maybe the mistake here is caused by the wrong tag associated to RADICAL which is treated as JJ rather than as NN. It is obvious that by using Stanford parser a certain level of error rate is expected: it would have been interesting to know what additional error rate is introduced by the conversion algorithm, but the evaluation is missing yet. It is also important to remember that Stanford parser only produces a surface level representation with some additional predicate argument completion for passive structures and some control infinitivals. So it is impossible to judge whether the reduction process – also in light of the example discussed above – has positively contributed to the final representation or not. Certainly the most important contribution, the elimination of free variables and the control exerted on the predicates arity, constitute by themselves already an important goal achieved. Of no real consequences is on the contrary the added feature regarding the insertion of the sense synset index directly in the overall logical form representation, the one delimited by ILF: this fact is disputable simply by iteself seen that there has been no word sense disambiguation of the gloss as a whole, something commented upon also by the authors in their conclusions (ibid.,35).

## 5. Conclusions

Eventually we may comment that there are a number of resources available with Logical Forms representations of WordNet glosses, and a number of algorithms which can be used off-the-shelf to produce Logical Forms from PTB constituency based phrase structure representations: none of these resources is however usable as is, do to error rates which average 30%. Improvements can be achieved by manual correction of all the LFs contained in these resources. This is an option that we intend to carry out in a local project that will be the followup of a MA degree thesis that started this research. The research has focussed on the typing of the mistakes present in the resource itself: this has been made easier by the fact that in both resources analysed, the conversion into LFs has started from the output of a syntactic parser – in the case of XWN, two constituency parsers, while in ILF, one dependency parser. The result of the manual corrections will be made available online to be accessed freely by anyone interested in using them.

## References

Agerri, R. and Anselmo Peñas (2010). On the Automatic Generation of Intermediate Logic Form for WordNet glosses, In *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (Cicling-2010),* LNCS Vol.

6008, Springer.

Alshawi, H., Pi-Chuan Chang, M. Ringgaard. (2011). Deterministic Statistical Mapping of Sentences to Underspecified Semantics, in Johan Bos and Stephen Pulman (editors), Proceedings of the 9th International Conference on Computational Semantics, IWCS,15-24.

Clark, P., Fellbaum, C., Hobbs, J.R., Harrison, P., Murray, W.R., Thompson, J.: Augmenting WordNet for Deep Understanding of Text. In: Bos, J., Delmonte, R. (eds.) *Proceedings of STEP 2008, Conference on Semantics in Text Processing. Research in Computational Semantics,* vol. 1, pp. 45–57. College Publications (2008)

Delmonte R., (2007). Computational Linguistic Text Processing – Logical Form, Semantic Interpretation, Discourse Relations and Question Answering, Nova Science Publishers, New York.

Yuqing Guo, HaifengWang, Josef van Genabith, Recovering Non-Local Dependencies for Chinese, 2007. In *Proceedings of the Joint Conference on EMNLP and Computational Natural Language Learning,* pp. 257–266.

Harabagiu, S.M., Miller, G.A., Moldovan, D.I.: *eXtended WordNet - A Morphologically and Semantically Enhanced Resourc*e (2003).
Available at http://xwn.hlt.utdallas.edu.

Information Science Institute, University of Southern California: Logical Forms for WordNet 3.0 glosses (2007), http://wordnetcode.princeton.edu/standoff-files/wn30-lfs.zip

Mihalcea, R., and Dan I. Moldovan, (2001). *eXtended WordNet: progress report,* In: Proceedings of NAACL Workshop on WordNet and Other Lexical Resources, Pittsburgh, 95-100.

Moldovan, D., Rus, V.: Explaining Answers with Extended WordNet. In: Proceedings of the Association for Computational Linguistics, ACL 2001 (2001)

Moldovan, D., S. Harabagiu, R. Girju, P. Morarescu, F. Lacatusu, A. Novischi, A. Badulescu, O. Bolohan, *LCC Tools for Question Answering,* In: Proceedings of TREC, 2002.

Moldovan, D., Vasile Rus: Transformation of WordNet Glosses into Logic Forms. FLAIRS Conference 2001: 459-463.

Moldovan, D., Vasile Rus: Logic Form Transformation of WordNet and its Applicability to Question Answering. ACL 2001: 394-401.

Ovchinnikova, E., N. Montazeri, T. Alexandrov, J. R. Hobbs, M. C. McCord, R. Mulkar-Mehta, (2011). Abductive Reasoning with a Large Knowledge Base for Discourse Processing, in J. Bos and S. Pulman (editors), In Proceedings of the 9th International Conference on Computational Semantics, IWCS, 2Oxford, pp. 25-234.

Ramsay, A.M. and D.G.Field (2008). Speech acts, epistemic planning and Grice's maxim. Journal of Logic and Computation 18(3), 431-457.

Rus, V. and Dan I. Moldovan: High Performance Logic Form Transformation. International Journal on Artificial Intelligence Tools 11(3): 437-454 (2002)

Rus, V., Dan I. Moldovan, Orest Bolohan, (2002). Bracketing Compound Nouns for Logic Form Derivation. FLAIRS Conference 2002: 198-202.

Rus, V., Alex Fit-Florea, (2004). A Model for Identifying the Underlying Logical Structure of Natural Language. PRICAI 2004: 977-978.

Rus, V., (2004). Experiments with Machine Learning for Logic Arguments Identification. MAICS 2004: 40-47.

Rus, Vasile, (2004). *A First Evaluation of Logic Form Identification Systems,* SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text.

Vasile Rus, *High Precision Logic Form Transformation,* 2001. In Proceedings of ICTAI'2001. pp.288-288.

WordNet Gloss Disambiguation Project, Princeton University: Semantically annotated gloss corpus (2008), http://wordnet.princeton. edu/ glosstag.shtml

# Interoperable Spatial and Temporal Annotation Schemes

**Kiyong Lee**

Korea University
Seoul 137-767, Korea
ikiyong@gmail.com, klee@korea.ac.kr

**Abstract**

ISO-TimeML (2012) was just published as an international standard for the annotation of temporal and event-related information in language. Almost at the same time, Pustejovsky and Moszkowicz (2012) produced a revised version of ISO-Space specifications as a spatial annotation scheme. The purpose of this paper is to argue for the need of making these two annotation schemes interoperable to allow a unified treatment of annotating spatial and temporal information in language. This task is mainly motivated by many occurrences of spatio-temporal signals (e.g., *at, in, after*) in text that trigger both spatial and temporal relations between various types of basic elements annotated to text offsets or segments, called *markables*. We argue that these two semantic annotation schemes can be made interoperable by merging some of their specifications, especially concerning the use of spatial or temporal signals and those relations triggered by these signals and, furthermore, that this merging results in designing an integrated spatio-temporal annotation and interpretation scheme.

## 1. Introduction

This paper concerns the interoperability of two semantic annotation schemes, ISO-TimeML (2012) and ISO-Space (Pustejovsky and Moszkowicz, 2012). ISO-TimeML is an international standard, published by ISO, for the annotation of temporal and event-related information in natural language, while ISO-Space is an emerging international standard for annotating spatial and spatio-temporal information in natural language that was proposed by Pustejovsky and Moszkowicz (2012). The purpose of this paper is to extend some of the specifications introduced in ISO-Space to the possible reformulation of ISO-TimeML or to make some changes in ISO-Space, thereby making them interoperable. This paper focuses on the use of signals, namely English prepositions such as *at*, *in*, *after*, *from*, and *to*. In the semantic annotation of a text, these signals trigger both spatial and temporal relations between some basic elements that are annotated to text offsets or segments, called *markables*, in the text. This is illustrated by the following example:

(1) Mia left home $at_{s1}$ two o'clock to drink tea $at_{s2}$ the August Moon Teahouse.

Here are two occurrences of the preposition *at*. They are both understood as locating a motion (*left home*) or an event (*drink coffee*) at some particular point in time (*two o'clock*) or space (*the August Moon Teahouse*).

These two signals are treated separately in two different annotation schemes, ISO-TimeML and ISO-Space, triggering two different relational links between annotated markables in text. The temporal use of $at_{s1}$ in the above example, for instance, triggers a Temporal Link (TLINK) in ISO-TimeML, whereas the spatial use of the same preposition $at_{s2}$ in the same example above triggers a Qualitative Spatial Link (QSLINK) in ISO-Space. We find sufficient evidence from the spatial and temporal uses of prepositions in English and other types of signals in other languages for supporting a unified treatment of such signals and the relations triggered by them, although we focus on the uses of English prepositions and Korean spatial signals in this paper and leave the general discussion of multilingual issues to other occasions (see Lee et al. (2011)).

Neither the conformance of ISO-TimeML to ISO-Space nor their interoperability, however, alters the basic framework of either of the annotation schemes. Each annotation scheme is considered as having a structure $< M, B, R, @ >$, where $M$ is a set of markables, $B$ a list of basic entities to be annotated to $M$, $R$ a list of (binary) relations over $B$, and @ is a set of attribute-value functions associated with each element in $B$ or each relation in $R$. What is to be most affected in our modifications mostly concerns @, the specifications of attributes and values, while the basic entities $B$ and the relations $R$ in each of the annotation schemes are almost or totally preserved.

The rest of the paper develops as follows: Section 2 The annotation Scheme of ISO-Space, Section 2 Spatio-temporal Uses of English Prepositions, Section 4 Spatial Signals in Korean, Section 5 Making ISO-Space and ISO-TimeML Interoperable, and Section 6 Concluding Remarks.

## 2. The Annotation Scheme of ISO-Space

The annotation scheme of ISO-Space consists of two components: a set of basic entities and a set of spatial relations over them. Basic spatial entities are of four types: (1) location types, (2) non-location types,(3) spatial signals, and (4) measure types.

Locations are annotated as either PLACE or PATH. Non-location elements are annotated as SPATIAL_NE (spatial named entity), MOTION, and EVENT (non-motion type). Spatial signals, annotated as SPATIAL_SIGNAL, are mostly prepositions in English. MEASUREs refer to distance and other dimensions such as length and volume.

There are four spatial relation tags in ISO-Space: (1) QSLINK, (2) OLINK, (3) MOVELINK, and (4) MLINK. QSLINK is for qualitative spatial link, OLINK provides information about orientations, MOVELINK involves motions, and the measure link MLINK defines the dimensions of a location.

For each of the four basic spatial entities and the four spatial relations, ISO-Space also specifies a list of attributes and a list of their possible values. The PLACE tag mostly inherits the attribute-value specification of SpatialML 3.0 (MITRE, 2009) and (Mani et al., 2010). The name *Boston*, for instance, can be annotated as below:

(2) place(pl1, type=PPL, ctv=CITY, form=NAM, state= MA, country=USA)

This annotation is understood as carrying the information that *Boston* with an **id** being **pl1** is a name of a populated place **ppl** of type **city** in the state of Massachusetts, USA. Basic spatial entities are directly tagged on text offsets or segments, called *markables*, in text. Spatial relations, on the other hand, relate elements annotated earlier in the text to other annotated elements. Consider an **id**-assigned text as below:

(3) [John$_{sne1}$] [lives$_{e1}$] [in$_{s1}$] [Boston$_{pl1}$].

Each of the square-bracketed markables with unique **id**'s can be annotated as below:

(4) spatial_entity(sne1, form=NAM)
event(e1, event_type=STATE, event_class= STATE)
spatial_signal(s1, cluster="in-1", semantic_type=topological)
place(pl1, type=PPL, ctv=CITY, form=NAM, state =MA, country=USA)

Here are two possible qualitative spatial links: one relates the spatial entity *John* to the place *Boston* and another relates the event *lives* to the place *Boston*. These links can be represented as below:

(5) a. qslink(qsl1, figure=sne1, ground=pl1, trigger=s1, relType=IN)

b. qslink(qsl1, figure=e1, ground=pl1, trigger=s1, relType=IN)

Consider another example:

(6) [John$_{sne1}$] [drove$_{m1}$] [from$_{s1}$] [Boston$_{pl1}$] [to$_{s2}$] [New York$_{pl2}$].

Each of the markables with unique **id**'s can also be annotated as below:

(7) spatial_entity(sne1, form=NAM)
motion(m1, motion_type=MANNER, motion_class= MOVE)
spatial_signal(s1, cluster=from-1, semantic_type=directional)
place(pl1, type=PPL, ctv=CITY, form=NAM, state= MA, country=USA)
spatial_signal(s2, cluster=to-1, semantic_type=directional)
place(pl2, type=PPL, ctv=CITY, form=NAM, state =NY, country=USA)

Then MOVELINK relates these annotations to each other.

(8) movelink(mvl1, trigger=m1, source=pl1, goal=pl2, mover=sne1, goal_reached=TRUE)

The other two links, OLINK (Orientatiol Link) and MLINK (Metric Link), can also similarly be introduced and discussed by referring to the current version of ISO-Space (Pustejovsky and Moszkowicz, 2012).

## 3. Spatio-temporal Uses of English Prepositions

Bennett (1975) analyzes 38 English prepositions[1] that are used either in a spatial sense or a temporal sense, or both. As he argues, most of them carry many different senses or uses that may be unified into core or more general meanings. Consider the entry of the preposition *at* in CCED (2006) (Collins COBUILD English Dictionary) that lists 19 senses. Out of 19, two senses are related to places, whereas two other senses are related to times. Here we cite four examples from the entry *at* in CCED (2006).

(9) a. We had a dinner **at** a restaurant in Attleborough...

b. I majored psychology **at** Hunter College.

c. The funeral will be carried out this afternoon **at** 3.00...

d. Bake emigrated to Australia with his family **at** 13...

Following Bennett (1975), we claim that these four senses can be combined into one meaning that refers to particular locations and that these locations can be interpreted with different senses, depending on the context of use. The preposition *at* is thus treated as having a meaning that can be characterized with the feature **locative**, while it can be interpreted as referring to diverse types of locations that may be either spatial or temporal, while locating some event or state at a particular point of places or times.[2]
Besides the preposition *at*, we find 12 prepositions that can be used in both spatial and temporal contexts: *after, around, before, by, from, in, into, on, past, through, to* and *towards*. Here are two simple examples from CCED (2006):

(10) a. after:
*After* breakfast Amy ordered a taxi... (time)
A few kilometers *after* the village, turn right to Mountelabate. (place)

b. before:
My husband rarely comes to be *before* 2 or 3am. (time)
They drove through a tall iron gate and stopped *before* a large white villa. (place)

Now consider the preposition *in*. Besides its entry as an adverb, there are four different entries of the preposition *in* in CCED (2006). Here are some examples:

(11) a. He was *in* his car. [artifact]

b. Don't stick too precisely to what it says *in* the book. [conceptual object]

c. ... that early spring day *in* April 1949... [time]

[1]This list includes one preposition, namely *in back of*, which he claims is not in his own dialect.

[2]The feature **locative** may also be interpreted as referring to a functional location, while allowing the interpretation of a location as a functional entity, as illustrated by *Hunter College*.

d. He walked two hundred and sixty miles *in* eight days. [amount of time]

e. Economy was *in* trouble... [state]

To interpret all these different uses of the preposition *in* in a uniform way, the notion of **locative** should be made transferable to various types of entities such as artifacts, conceptual objects, times, amount of time, and states.

In ISO-Space there are two semantic types of spatial signals: topological and directional. Extended to temporal entities, the semantic type **topological** can be replaced by **locative**. This will differentiate two different uses of the preposition *in* as in the following:[3]

(12) a. My parents live [$\text{in}_{locative}$] New Zealand now.

b. I never went [$\text{in}_{directional}$] pubs.

Both *at* and *in* may be treated **locative** signals, while the preposition *in* is also treated as **directional** as used in (b) above.

The prepositions such as *from* and *to* are used to indicate the beginning point and the end point of events both in a spatial context or in a temporal context, as shown below:

(13) a. Spatial context: John walked *from* Boston *to* Cambridge.

b. Temporal context: John worked on a farm for twelve hours *from* six in the morning *to* six in the afternoon.

The annotation of these two sentences is thus expected to follow a unified format.

We then find cases where the distance from one place to another is given either in spatial terms or in temporal terms. Here is an example taken from the web:

(14) Distance from Seoul to Chiang Mai is: 2156.8 miles (3471.1 kilometers / 1873 nautical miles). Approximate flight duration time from Seoul to Chiang Mai is 4 hrs, 29 mins.

These pieces of evidence in the use of language, especially that of prepositions in English,[4] strongly argue for the necessity of integrating spatial and temporal annotations into a unified annotation scheme or making them interoperable.

## 4. Spatial Signals in Korean

There are two constructional types of spatial signals in Korean. One is a simple type that consists of a single particle (e.g., *ey* (locative) 'at') or sometimes consists of a sequence of basic particles (e.g., *-eyse-pwute* (source, path start) 'from') and another is a complex type that consists of a noun followed by a simple type spatial signal (e.g., *wi-ey* (locative) 'on'/'at the upper part of'), where the noun *wi* refers to some upper part of a location. In this section, we aim to show how the specification of spatial signals in ISO-Space (Pustejovsky and Moszkowicz, 2012) apply to these two types of spatial signals in Korean.

[3]These examples are taken from LDOCE5 (2009).

[4]Strong evidence is shown in other languages such as Korean, Japanese, and Chinese. See Lee et al. (2011) and Sohn (1999).

### 4.1. Simple Type Spatial Signals

ISO-Space proposes two semantic types of spatial signals: **topological** and **directional**. The terms **locative** and **directional** are, however, well-established grammatical terms, especially for the categorization of case marking particles in Korean. (See Lee (1999) and Sohn (1999)). Furthermore, as stated earlier, the term **locative** applies not only to English prepositions or Korean particles as triggering topological or spatial relations, but also temporal and other types of relations between annotated elements in language. We thus propose that the term **topological** be replaced by the term **locative** to cover both spatial and temporal uses of relational signals in the semantic annotation of language as well as to conform to the established terminology in grammar.

The function of English prepositions as spatial signals is taken up by nominal particles in Korean. As an agglutinative language, Korean has over 100 basic particles that are suffixed to nouns (e.g., $seoul_{noun}$-$ey_{particle}$ *'at/in Seoul'*).[5] It is also possible to generate around 3,000 particles by combining these basic particles (e.g., $seoul_{noun}$-$ey_{particle}$-$nun_{particle}$). (See Lee (1999).) Out of those over 100 basic particles, there are only a few basic particles that function as spatial signals.

In this section, we first discuss two **locative** particles *-ey* and *-eyse* and then one **directional** particle *-(u)lo* to support the two-way distinction of semantic types of spatial signals into **locative** and **directional**. Thirdly, we introduce two special particles *pwute* ('from') and *kkaci* ('to') that mark the start and the end point of a path, respectively.

#### 4.1.1. Locative Particles

The particles *-ey* and *-eyse* are typical locatives in Korean. Consider the following examples:[6]

(15) a. -ey 'at' (locative):
mia-nun cip-*ey* issta
Mia-TOP home-LOC is
'Mia is at home'

b. -eyse 'at' (locative):
mia-nun cip-*eyse* swiessta
Mia-TOP home-LOC rested
'Mia rested at home'

Sohn (1999) differentiates the two locative particles by naming *-ey* a stative locative and *-eyse* a dynamic locative. Consider:

(16) a. state:
haksayngtul-i motwu kyosil-*ey* namassta
students-NOM all classroom-LOC remained
'The students all remained in the classroom'

[5]Although they are most frequently suffixed to nouns, these particles, especially so-called **special particles**, may be suffixed to other categories than nouns.

[6]TOP stands for the topic marker, LOC for the locative marker, NOM for the nominative marker, and DIR for the directional marker.

b. dynamic:
haksayngtul-i motwu kyosil-*eyse* kongpuhayssta
students-NOM all classroom-LOC studied
'The students all studied in the classroom'

These examples support Sohn (1999)'s claim.
As Lee and Chae (1999) point out, it is certain that the choice between *-ey* and *-eyse* depends on the type of a verb that is used with either of them, but it is difficult to identify exactly what that type is. Consider the following:

(17) a. state or activity?:
cwi-ka ce kwumeng-*ey* swumessta
rat-NOM that hole-LOC hid (itself)
'The rat hid (itself) in that hole'

b. state:
cwi-ka ce kwumeng-*eyse* cwukessta
rat-NOM that hole-LOC died
'The rat died in that hole'

Here, both of the verbs *swumta* 'to hide' and *cwukta* 'to die' refer to non-dynamic states. The former may also be interpreted as referring to an activity, while the latter is interpreted as referring to a state. Neither of these interpretations supports Sohn (1999)'s classification of locatives into stative and dynamic locatives.
Consider the verb *salta* 'to live'. It is a stative verb, but can have either the construction PLACE-*ey* or the construciton PLACE-*eyse* as its complement. Here are examples:

(18) mia-nun pusan-*ey/eyse* salassta
Mia-TOP Busan-LOC lived
'Mia lived in Busan'

Hence, Sohn (1999)'s position does not hold here.
Nevertheless, this issue creates no problem for the annotation of text in general, for we are not dealing with generation issues.
There is, however, one case that concerns the annotation of these locatives *-ey* and *-eyse* as spatial signals triggering the semantic roles of @goal and @source, respectively. When used with a motion verb, the particle *-ey* signals the goal or end point of that motion referred by the verb, while the particle *-eyse* signals the source or start point of the motion.

(19) a. mia-ka seoul-ey wassta
Mia-NOM Seoul-LOC:GOAL came
'Mia came to Seoul' (goal, end point)

b. mia-ka pusan-eyse wassta
Mia-NOM Busan-LOC:SOURCE came
'Mia came from Busan' (source, start point)

Hence these particles also have a directional interpretation, but such an interpretation is only possible when these locatives are used with a motion verb.

#### 4.1.2. Directional Particles

The typical directional particle is *(u)lo* that relates a place to a motion as its goal, as illustrated below:

(20) a. mia-ka mikuk-*ulo* kassta
Mia-NOM USA-DIR, GOAL went
'Mia went to USA'

b. yong-i nyuyok-*ulo* ttenassta
Yong-NOM New York-DIR, GOAL left
'Yong left for New York'

Here, the place *mikuk* 'USA' is the goal of the motion *kassta* 'went' and the place *nyuyok* 'New York' the goal of the motion *ttenassta* 'left'.
As noted earlier, the locative *-ey* can also be used in a directional sense by relating a motion to a place as its goal. There is, however, a basic difference between the directional use of *-ey* and that of *-(u)lo*, as illustrated by the following pairs of examples:

(21) a. mia-ka mikuk-*ey* tochakhayssta
Mia-NOM USA-LOC:DIR, GOAL arrived
'Mia arrived in America'

b. *mia-ka mikuk-*ey* ttenasst
Mia-NOM USA-LOC:DIR, GOAL left
UNACCEPTABLE

(22) a. mia-ka mikuk-*ulo* ttenassta
Mia-NOM USA-DIR, GOAL left
'Mia left for USA'

b. *mia-ka mikuk-*ulo* tochakhayssta
Mia-NOM USA-DIR arrived
UNACCEPTABLE

This indicates that the particle *-ey* can be used as a directional signal only if the goal is reached, whereas the particle *-ulo* can be used with a motion verb without implicating its accomplishment. The attribute @goal_reached for the movement link MOVELINK in ISO-Space thus plays a significant role here. In (a) of the first examples with the particle *-ey*, the value of the attribute @goal_reached should be TRUE. As shown by the second pair of examples above with the particle *-ulo*, the particle *-ulo* cannot be used with a motion whose goal has been reached.

#### 4.1.3. Path-related Particles

There are two types of path-related particles in Korean. One type indicates the start or source of a path and another the end or goal of a path. The particles *-pwute*, *-eyse*, *-eyse-pwute*, *-ulo-pwute* all belong to the first type. The particle *kkaci*, on the other hand, belongs to the second type. These two types are often used as pairs, indicating the start and the end of a path that is created by a motion. Here are examples.

(23) a. yong-un cip-eyese-pwute hakkyo-kkaci kelessta
Yong-TOP home-START, SOURCE school-END, GOAL walked
'Yong walked from home to school'

b. yong-un mayil hakkyo-kkaci kele kassta
Yong-TOP everyday school-END, GOAL walking went
'Yong walked to school everyday'

Associated with MOVELINK in ISO-Space, two places *cip* 'home' and *hakkyo* 'school' may be interpreted as being the source and the goal of the motion *kelessta* 'walked'.
Consider, however, a more complex example like the following:

(24) yong-un mayil cip-eyese-pwute hakkyo-kkaci 10 mail-ul kele kassta
Yong-TOP everyday home-START, SOURCE school-END, GOAL 10 miles walked
'Yong walked 10 miles from home to school everyday'

Here, the path *cip-eyese-pwute* 'from home' *hakkyo-kkaci* 'to school' can be linked to the distance *10 mail* '10 miles', with *cip* 'home' being the start point and *hakkyo* being the end point of the distance. Hence, the spatial annotation of the example above involves both MOVELINK (movement link) and MLINK (metric link) in ISO-Space.
Furthermore, these path or distance-related particles can be used without any motion being mentioned. Here is an example:

(25) cip-eyese-pwute hakkyo-kkaci-nun 10 mail-ita
home-START school-END-TOP 10 miles-COPULA
'from home to school is 10 miles'

These particles can also be used without referring to any specific distance.

(26) yeki-se-pwute ceki-kkaci-nun nay ttang-ita
here-START there-end-TOP my land-COPULA
'From here to there is my land'

Some boundary or dimension of the land is implicitly assumed, allowing the use of MLINK in ISO-Space.

### 4.2. Complex Spatial Signals

Associated with the attribute @RelType of OLINK (Orientation Link), the following values are introduced:

(27) NEAR, ABOVE, BELOW, FRONT, BEHIND, LEFT, RIGHT, NEXT TO, NORTH, ...

In Korean and possibly in Japanese and Chinese, these values are expressed by nouns that refer to particular parts of a location. Here are examples:

(28) a. vertical: *wi* 'above', *alay* 'below', *mit* 'under',

b. horizontal: *aph* 'front', *twi* 'behind', *yeph* 'side, next to', *oyn-ccok* 'left', *olun-cchok* 'right'

c. azimuth(?): *tong, tong-ccok* 'east', *se, se-ccok*, *nam, nam-ccok* 'south', *pwuk, pwuk-ccok* 'north',...

In order to function as spatial signals, these orientation nouns each combine with one of the particles that function as spatial signals. Only when combined with a particular spatial particle, these nouns become either locative or directional, or either source or goal. Here are examples:

(29) a. chayk-i chayksang wi-ey issta
book-NOM desk above-LOC is
'The book is on the desk

b. sicheng aph-eyse mannaca
city hall front-LOC let's meet
'Let's meet in front of the city hall'

c. wuli-nun namccok-ulo kele kassta
we-TOP south-DIR walk went
'We walked to the south'

These orientation nouns refer to certain parts of a location without specifying their projective extent. Consider the following examples:

(30) a. pihayngki-ka pata wi-lul nalassta
airplane-NOM sea above-ACC[7] flew
'An airplane flew over the sea'

b. pay-ka pata wi-lul tallyessta
boat-NOM sea above-ACC ran
'A boat sailed/ran over the sea'

Neither of the sentences here specifies how far above from the sea each of these events took place. Such a specification is made only through the understanding of each of the actual circumstances. Understanding these circumstances, the intelligent annotator may be able to specify such projective extents. For this specification, Pustejovsky and Moszkowicz (2012) provides the attribute @projective as well as other related attributes, @figure, @ground, @frametype, and @referencePt with their possible values.

(31) a. an 'inside', pakk 'outside', ka 'around', kunche 'near place'

b. three dimensional: sok 'inside'

The specification of RCC8+ (the region connection calculus) as values of the attribute @relType of QSLINK (Qualitative Spatial Link)[8] may not fully depend on these spatial signals either, but again on our understanding of the whole circumstances associated with the type of each particular event. This again is a very demanding task for the annotator, but may be necessary for the useful applications of semantic annotation in general.

## 5. Making ISO-Space and ISO-TimeML Interoperable

Obviously the current versions of ISO-TimeML and ISO-Space differ in the mode of representing their annotations. Annotated spatial and temporal information, however, show isomorphic resemblance in anchoring, orientation, and measure and in semantic interpretation in general. In this section we will show how this isomorphic resemblance can be captured, thus making the two annotation schemes interoperable.

[7] ACC stands for the **accusative** case

[8] RCC8+ has 9 possible values such as DC (Disconnected), EC (External Connection), PO (Partial Overlap), etc. See MITRE (2009) and ISO-TimeML (2012) for details.

### 5.1. Representing Annotations

As its name implies, ISO-TimeML as a revised version of TimeML simply adopted XML to represent its annotations. Instead of XML, the current version of ISO-Space uses a predicate-logic-like format for the representation of its annotations. Here are examples:

(32) a. Mia left$_{e1}$ for$_{s1}$ Busan$_{pl1}$ yesterday$_{t1}$.

b. ISO-TimeML: <TLINK eventID="e1" relatedToTime="t1" relType="DURING"/>

c. ISO-Space: movelink(mvl1, goal=pl1, trigger=s1, goal_reached=false)

In principle, the choice of a particular representation format should not affect the basic framework of annotation schemes. The information content of annotation should also remain the same independent of how it is represented. One possible exception, however, might arise when we try to group some relational links under one representation frame.[9] Since there is no such grouping of links in the current version of ISO-TimeML, we assume that all of the XML representations in ISO-TimeML remain equivalent even if they are converted to predicate-logic-like formats as in ISO-Space.

### 5.2. Spatial and Temporal Anchoring of Events

Events may be anchored to a place or a time. Here is an example from CCED (2006):

(33) a. raw text: Mary Martin has died at her home in California at the age of seventy-six.

b. Spatial markables: Mary Martin [has died$_{e1}$] [at$_{s1}$] [her home$_{pl1}$] [in$_{s2}$] [California$_{pl2}$] at the age of seventy-six.

c. Temporal markables: Mary Martin [has died$_{e1}$] at her home in California [at$_{s3}$] [the age of seventy-six$_{t1}$].

(a) is a raw text. All of the spatial basic elements are marked up in (b), while all of the temporal basic elements are marked up in (c). Then here are two anchoring relations: one is a spatial anchoring that locates the event **e1** of Mary Martin's death at a place **pl1** (*her home*), while the other relation is a temporal anchoring that locates the same event at a time **t1** (*the age of seventy-six*).

As the spatial signal **s1** triggers QSLINK, this spatial anchoring is represented by ISO-Space like the following:

(34) qslink(qsl1, figure=e1, ground=pl1, trigger=s1, relType=IN)

This can easily be interpreted as stating that the event of Mary Martin's death occurred at the place **pl1**.

The signal **s3** (*at*), however, triggers a temporal relation. So we have to follow ISO-TimeML to annotate such a relation and get the following:

[9] This representation issue will be discussed in a later paper.

(35) <TLINK eventType="e1" relatedToTime="t1" signalID="s3" relType="IN"/>

To make this representation comparable to that of ISO-Space, we introduce two modifications to ISO-TimeML. First, instead of an XML representation, we adopt a predicate-logic-like language to represent temporal annotations in ISO-TimeML as in ISO-Space.[10] Second, we modify the attribute names @eventType, @relatedToTime and @signalID to more general names @figure, @ground, and @trigger, respectively. With these two modifications, we have the following:

(36) tlink(tl1, figure=e1, ground=t1, trigger=s3, relType=IN)

This is then interpreted as stating that the event **e1** of Mary Martin's death occurred at the time **t1** (*the age of seventy-six*) in the same manner of interpreting the **qslink(qs1)**. This means that one single rule guarantees the interpretation of both spatial and temporal links.

### 5.3. Spatial and Temporal Orientations

There are other types of temporal relations such as the precedence relation than the anchoring of an event to a time. The precedence relation can be expressed by the prepositions *before* and *after*, as in the following example again from CCED (2006):

(37) a. My husband rarely comes$_{e1}$ to bed before$_{s1}$ [2 or 3am$_{t1}$].

b. After$_{s3}$ breakfast$_{e2}$ Amy ordered$_{e3}$ a taxi...

Adopting the predicate-logic-like representation format, ISO-TimeML annotates the above texts as below:

(38) a. tlink(tl1, eventID=e1, relatedToTime=t1, signalID=s1, relType=BEFORE)

b. tlink(tl1, eventID=e3, relatedToEvent=e2, signalID=s2, relType=AFTER)

These relations cannot be related to QSLINK, but rather to Orientation Link (OLINK) in ISO-Space, for the values of @relType like BEFORE and AFTER are not parts of the region connection calculus, called *RCC8+*. Note that this calculus only treats spatial relations, but of those in OLINK.[11]

As discussed in section 3, the English prepositions *before* and *after* are used both in a temporal context and in a spatial context. Consider the following partially marked-up example:

(39) The Governor appeared$_{e1}$ before$_{s1}$ the committee$_{sne1}$ before$_{s2}$ noon$_{t1}$.

In ISO-Space, the spatial relation between the event **e1** (*appeared*) and the spatial named entity **sne1** (*the committee*) is captured by OLINK, whereas the temporal relation between the event **e1** and the time **t1** (*noon*) is captured by TLINK, as shown below:

[10] Now the modified ISO-TimeML should be named **ISO-Time**.

[11] Bennett (1975)(page 119) also discusses the spatial use of the preposition *before*, citing examples like *She set an enormous meal before him, He was ordered to appear before the magistrate*.

(40) a. Spatial Orientation:
olink(ol1, figure=e1, ground=sne1, trigger=s1, relType=FRONT)

b. Temporal Orientation:
tlink(tl1, eventID=e1, relatedToTime=t1, signalID=s2, relType=BEFORE)

Here we can make three modifications. First, the attribute names @eventID and @relatedToTime can also be generalized to the attribute names @figure and @ground, respectively, as are generalized for the spatial and temporal anchoring.
Second, following Bennett (1975), the attribute values FRONT and BEFORE can be unified to a more general name ANTERIOR, while the value names BEHIND and AFTER are unified to the name POSTERIOR.
Third, we can have OLINK to take over the function of TLINK related to the temporal precedence relation, for this can be viewed as a type of orientation relation. The third modification, however, requires the introduction of OLINK into ISO-TimeML, thus changing the annotation scheme of ISO-TimeML. We then have the following unified representations:

(41) a. Spatial Orientation (revised):
olink(ol1, figure=e1, ground=sne1, trigger=s1, relType=ANTERIOR)

b. Temporal Orientation (revised):
olink(tl1, figure=e1, ground=t1, trigger=s2, relType=ANTERIOR)

### 5.4. Paths and Durations

Paths and durations have almost the identical construction possibly with the same interpretation structure. Both of them, for instance, may be expressed with the use of prepositions like *from* and *to*. Consider the following examples involving the motion of walking:[12]

(42) a. Path: John$_{sne1}$ walked$_{m1}$ from$_{s1}$ Boston$_{pl1}$ to$_{s2}$ Cambridge$_{pl2}$.

b. Duration: John$_{sne1}$ walked$_{m1}$ from$_{s3}$ two$_{t1}$ to$_{s4}$ four$_{t2}$.

In ISO-Space, example (a) is annotated with MOVELINK (Movement Link) as below:

(43) movelink(mv1, trigger=m1, source=pl1, goal=pl2, mover=snel, goal_reached=TRUE)

Since the motion of John's walking is understood as going through some path *from Boston to Cambridge*, it can also be accompanied by a PATH annotation:

(44) path(p1, beginPoint=pl1, endPoint=pl2)

On the basis of the PATH annotation, the annotation MOVELINK may refer to that path, as shown below:[13]

(45) movelink(mv1, trigger=m1, pathID=p1, mover=snel, endPoint_reached=TRUE).

Note that the attribute @pathID is necessary to handle cases like the following:

(46) a. John drove through Route 66.

b. John$_{sne2}$ drove$_{m2}$ through$_{s2}$ [Route 66]$_{p2}$.

c. movelink(mv2, trigger=m2, pathID=p2, move= sne2, goal_ reached=FALSE)

Now consider the example given above that involves a (temporal) duration

(47) a. John walked from two to four.

b. John$_{sne1}$ walked$_{m1}$ from$_{s3}$ two$_{t1}$ to$_{s4}$ four$_{t2}$.

As it is, ISO-Space cannot annotate this sentence, for its source and goal are not spatial entities. ISO-TimeML, however, annotates it, as shown below:[14]

(48) a. timex3(t1, type=TIME, value=T14:00)
timex3(t2, type=TIME, value=T16:00)
timex3(t3, type=DURATION, value=P2H, beginPoint =t1, endPoint=t2, temporalFunction=TRUE)

b. tlink(timeID=t1, signalID=s3, relatedToTime=t3, relType=BEGINS)
tlink(timeID=t2, signalID=s4, relatedToTime=t3, relType=ENDS)
tlink(eventID=m1, relatedToTime=t3, relType= SIMULTANEOUS)

Some modifications need be made on these annotations. First, the element TIMEX3 that deals with durations can be differentiated with other uses of TIMEX3 just as the element PATH is differentiated from the element PLACE in ISO-Space. This can be done by introducing a new element DURATION comparable to PATH and also by generalizing the element MEASURE in ISO-Space to temporal measure. Here is an example:

(49) a. timex3(t3, type=DURATION, value=P2H, beginPoint =t1, endPoint=t2, temporalFunction=TRUE)

b. duration(d1, beginPoint =t1, endPoint=t2, value= me1)
measure(me1, value=2, unit=hour)[15]

Second, ISO-Space fails to make explicit use of the spatial signals such as *from* and *to* as triggers of the attributes @beginPoint and @endPoint or the attributes @source and @goal. The specification of the element QSLINK may be modified for this purpose, as illustrated below:

(50) a. John$_{sne1}$ walked$_{m1}$ from$_{s1}$ Boston$_{pl1}$ to$_{s2}$ Cambridge$_{pl2}$.

[12]Example (a) is from Pustejovsky and Moszkowicz (2012).
[13]The attribute name @goal_reached is replaced by the name @endPoint_reached.
[14]See ISO-TimeML (2012), p. 19.
[15]Here, the attribute name @value may be replaced by @quantity.

b. path(p1, beginPoint=pl1, endPoint=pl2)

c. qslink(qsl1, figure=pl1, ground=p1, trigger=s1, relType=BEGINS)
qslink(qsl2, figure=pl2, ground=p1, trigger=s2, relType=ENDS)

This modification requires the addition of BEGINS and ENDS to the values of @relType for QSLINK in ISO-Space. As noted earlier, these QSLINKs are independent of the motion itself and simply specify the beginning and end point of a path.[16]

### 5.5. Measurements: Distance and Time Amount

Distances are expressed in spatial terms or temporal terms. Consider the following information obtained from the web:

(51) Distance from Seoul to Chiang Mai is: 2156.8 miles (3471.1 km/ 1873 nautical miles). Approximate flight duration time from Seoul to Chiang Mai is 4 hrs, 29 mins.

By modifying the specification of MLINK, ISO-Space can annotate the information about distance as shown below:

(52) measure(m1, value=2156.8, unit=mile)
path(p1, beginPoint=Seoul, endPoint=Chaing_Mai)
mlink(ml1, val=m1, pathID=p1, relType=DISTANCE)

The flight duration time can also be annotated as below:

(53) measure(m2, value={4:29} unit={hour:minute})[17]
path(p1, beginPoint=Seoul, endPoint=Chaing_Mai)
mlink(ml1, val=m2, pathID=p1, relType=DURATION)

## 6. Concluding Remarks

Other attempts have been made to integrate temporal annotation and spatial annotations into one unified format. One recent report was presented by Schuurman and Vandeghinste (2011), introducing the spatiotemporal annotation schema STEx. But no concrete proposal has been made concerning the interoperability of spatial and temporal annotation schemes, although various issues of interoperability and conformance have been discussed at the level of annotation, motivating workshops such as our ISA workshop. The obvious reason is that ISO-TimeML (2012) was just published, while the specifications of ISO-Space (Pustejovsky and Moszkowicz, 2012) keep being revised.

This paper touched on various aspects of these two annotation schemes. It has attempted to show sufficient ground to merge them into a unified annotation scheme or to make them interoperable mainly because the ontological and linguistic structures of space and time are very closely related, especially as shown by the occurrences of spatial and temporal signals, namely spatial and temporal uses of prepositions in English and spatial signals in Korean. Extended to languages other than English and Korean, there is expected to be more convincing evidence to argue for a unified treatment of the annotation of spatial and temporal information in natural language. This task is, however, left for the future work.

We argued for a unified treatment of spatial and temporal annotations mainly on the basis of use evidence in language. Because of a certain degree of isomorphism between space and time in nature or ontology, we also argue for such a treatment that results in interoperable semantic annotation and interpretation. This task again requires a more formal work for the future.

## 7. Acknowledgements

I owe thanks to Suk-Jin Chang, Jae-Woong Choe, and Roland H. Hausser for reading the pre-final version of this paper.

## 8. References

D. Bennett. 1975. *Spatial and Temporal Uses of English Prepositions: An Essay in Stratificational Semantics.* Longman, London.

CCED. 2006. *Collins COBUILD Adanced Learner's English Dictionary.* 6th edition. HarperCollins Publishers.

ISO/TC 37/SC 4/WG 2. 2012. *ISO 24617-1:2012(E) Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML).* The International Organization for Standardization, Geneva.

I. Lee and W. Chae. 1999. Lectures on Korean Grammar [written in Korean]. Hakyeon-sa. Seoul.

K. Lee. 1999. Computational Morphology [written in Korean]. Korea University Press. Seoul.

K. Lee, A. Fang, and J. Pustojvosky. 2011. Multilingual Verification of the Annotation Scheme ISO-Space. The First Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-1), the 5th IEEE International Conference on Semantic Computing. Stanford University, Palo Alto, CA.

MITRE. 2009. Annotation Scheme for Marking Spatial Expressions in Natural Language. Version 3.0. The MITRE Corporation.

I. Mani, C. Doran, D. Harris, J. Hitzeman, R. Quimby, J. Richer, B. Wellner, S. Mardis, and S. Clancy. 2010. SpatialML: Annotation scheme, resources and evaluation. *Language Resources and Evaluation*, 44:263-280.

J. Pustejovsky and J. Moszkowicz. 2012. ISO-Space Specifications: Version 4.1. Unpublished.

R. Quirk, D. Crystal, G. Leech, J. Wells, and D. Summers. 2009. *Longman Dictionary of Contemporary English for Advanced Learners.* 5th edition. Pearson Education Limited, Edinburgh.

I. Schuurman and V. Vandeghinste. 2011. Spatiotemporal annotation: interaction between standards and other formats. The First Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-1), the 5th IEEE International Conference on Semantic Computing. Stanford University, Palo Alto, CA.

H. Sohn. 1999. *The Korean Language*. Cambridge University Press. Cambridge.

[16]There is a certain amount of information overlap between the element PATH and the link QSLINK.

[17]The bracketing is tentatively introduced to represent the time amount like 2 hours 29 minutes.

# An XML Annotation Scheme for Space in an Italian Corpus

**Linda Meini, Giovanna Marotta, Alessandro Lenci, Margherita Donati**
Department of Linguistics, University of Pisa
Via Santa Maria 36, 56126 Pisa (Italy)
linda.meini@for.unipi.it, gmarotta@ling.unipi.it, alessandro.lenci@ling.unipi.it, margherita-donati@libero.it

**Abstract**

The new resource we present consists of a corpus of oral spatial descriptions performed by congenital blind and sighted Italian subjects. The collection of the data is part of a wider project on semantic representations in the language of the blind, carried out at the Department of Linguistics, University of Pisa. The long term goal of the project is to use the evidence collected on congenital blind subjects to get at a better understanding of the relationship between linguistic and perceptive information. The corpus is currently being enhanced with different layers of annotation, focusing on spatial information. The annotation allows us to highlight the effect of the specific lexical and grammatical features of Italian on the encoding of space (e.g. with respect to the way spatial relations are encoded in motion verbs). Our resource is not only one of the few annotated corpora of spoken Italian, but it is also the first one that focuses on spatial categories.

## 1. Introduction

Space has a fundamental role in human thinking and reasoning. Like time, it is clearly a core domain of human cognition, hence it represents an ideal testing ground for an in-depth analysis of the dynamic interplay between language and non-linguistic cognition. The relationship between cognitive representations and external reality is not trivial and it gets even more complex when spatial linguistic categories are factored in (for an introduction, see Marotta, 2010).

In linguistics, a longstanding debate exists between at least two main alternative models of the relationship between language and concepts.

According to a '*nativistic*' approach, the structures of spatial language are determined by our pre-linguistic categorization of space. The idea is that there is a restricted list of primitive, universal and innate notions, shared by all human beings; these notions include mostly topological relations (i.e. containment, support, contact, and proximity), and are mapped more or less directly into adpositions (Piaget and Inhelder, 1948; Miller and Johnson-Laird, 1976; Slobin, 1985; Talmy, 2000; Jackendoff, 2002).

However, scholars who analyzed spatial reference in a cross-cultural and cross-linguistic perspective have shown that both the kind of spatial relations encoded in language and the grammatical classes encoding spatial relations can vary dramatically from what we find in Western languages. These studies raised the question of how such linguistic variety can be found and accounted for if all human beings start with the same set of primitives. A new '*relativistic*' approach has been developed. According to various scholars spatial language is conditioned in several ways and to several degrees by cultural conventions, and reflects representations created by exposure to spatial words relating to one's native language. In brief, the structure and the lexicon of spatial language constrain the shape and the categories of "spatial thought" (Levinson 2003; Levinson and Wilkins, 2006; Landau *et al.* 2010).

The relationship between linguistic and non-linguistic categories grows in complexity when the reality 'out-there' is taken into account. Cognitive approaches assume that meanings coded by human language reflect reality as it is experienced by human beings. That is, our representation of reality is mediated by both the sensorimotor abilities of our bodies and the mental processes (basically automatic and unconscious) that organize perceptive stimuli. Therefore, some scholars (e.g. Talmy, 1983; Herskovits, 1986; Vandeloise, 1991) claimed that entities involved in spatial descriptions are not real objects, but rather geometrical abstractions of real entities that speakers conceptualize as points, lines, surfaces, or volumes. In addition, according to Vandeloise (1991) these geometrical abstractions are associated with prototypical functions that reflect how objects act in the world out there, and how we interact with them: for instance, a bowl is conceptualized as a volume with a containment function, which is coded in English by the preposition *in* 'in'. As for dynamic descriptions, Talmy (1983) has proposed various primitive templates or 'schemas' for representing motion. For instance, a moving object may be described as a geometric point moving along a path - that is a line - and/or towards another object conceptualized as a point: e.g. *the ball rolled along the ledge* (*toward the lamp*).

The project currently being developed at the Department of Linguistics, University of Pisa[1], aims at carrying out a comparative analysis of semantic representations in congenital blind subjects with respect to those of sighted subjects. The language and the conceptual structures of blind subjects have an inherently different experiential base, which is not grounded on the visual modality. Therefore, the semantic analysis of the language of congenital blinds can provide new insights on the important relationship between conceptual structures and sensory-motor information, and more in general on the relationship between language and experience.

[1] The project we are referring to is the PRIN project 2008-2010, number 2008CM9MY3.

## 2. Linguistic Specifications

The relationship between physical and spatial properties of the world 'out there', human cognition, and language is very complex. Landau *et al.* (2010) have recently emphasized that language is inherently selective, encoding certain distinctions and not others; this property enables language both to modulate attention and to serve as a mental pointer, indicating which of many possible representations we have in mind.

We would like to point out that two types of selectivity can be recognized: one depending on the speaker and the other depending on the structure of language. The first type is related to what the speaker wants to communicate and how (s)he conceptualizes a given scene in a given moment. In his studies on spatial language, Talmy (1983: 225) introduces the notion of 'schematization', i.e. "a process that involves the systematic selection of certain aspects of a referent scene to represent the whole, while disregarding the remaining aspects". Within a functionalist approach, some years later, Tyler and Evans (2003: 53) proposed the notion of 'vantage point', that "suggests that how a particular spatial scene is viewed will in large part determine the functional nature of a particular spatial scene". In other words, spatial relations between entities are not fixed once and for all, rather they largely depend on the speaker's perspective (Vandeloise, 1991: 23). Therefore, the linguistic description of a spatial scene is shaped by the specific speaker's point of view and his/her communicative purposes.

The second type of selectivity is related to the structure of language. Spatial relations are usually encoded by some grammatically defined classes:

- verbs of position and motion: e.g. Eng. *lie*, *sit*, *stand*, *roll*, *arrive*, *reach*, *go*;
- adpositions and particles or adverbs: e.g. Eng. *in*, *on*, *at*, *across*, *up*.

Languages typically lexicalize in each grammatical class specific semantic content, which varies from language to language.

For instance, languages differ with respect to which semantic components are lexicalized in the class of motion verbs (Talmy, 1985; Slobin, 2004). In English verbs typically lexicalize the conceptual components of Motion and Manner (e.g. *to roll*, *slid*, *walk*, *run*), while Path is encoded out of the verb by prepositions or particles (e.g. *to rolled off*, *walk into*, *go up*). By contrast, Italian verbs mostly lexicalize the conceptual components of Motion and Path (e.g. *uscire* 'to go out', *entrare* 'to go in', *salire* 'to go up/get on'), while Manner is optionally encoded out of the verb by adverbials or gerundive type constituents (e.g. *Sei salito a piedi?* lit. 'did you go up (on foot)', i.e. 'did you walk up?')[2]. There are other possibilities. Languages like Atsugewi (a Californian Indian language, now extinct) have a whole series of verb roots that lexicalize Motion and various kind of objects or materials as moving and located: e.g. *-lup-* 'for a small shiny spherical object to move/be-located', *-qput-* 'for loose dry dirt to move/be-located' (examples from Talmy, 1985). Languages can also differ with respect to which spatial relations are encoded by prepositions. For instance, Italian *su* can encode all the relations that in English are express by *on* (*upon*/*onto*), *over*, *above*, *on top of*, *up*. In fact, it is well known that spatial prepositions are ambiguous and highly context dependent (Vandeloise, 1991; Tyler and Evans, 2003; Meini, 2009). Another way to encode spatial relations is found, again, in Atsugewi. It has a set of verbal suffixes that encodes the Path and the type of objects or materials where the motion is directed to: e.g. *-ičt* 'into a liquid', *-mič* 'down into (/onto) the ground' (examples from Talmy, 1983).

The previous examples showed clearly that language's spatial system imposes a fixed form of structure on virtually every spatial scene. In other words, speakers cannot describe a spatial scene in just any way they might wish, rather they must choose among the word classes available in the organization of the lexicon of their specific language.

## 3. Data Collection

Within our project, 22 congenital blind subjects were selected, 10 females and 12 males, ranging from 21 to 72 years old (female average age: 47; male average age: 45). Of these 22 subjects 12 are from Tuscany, 5 from Liguria and 5 from Sardinia. On the basis of their personal data (such as age, gender, city of residence), as well as socio-linguistics parameters[3], 22 corresponding sighted subjects have been selected, with characteristics similar to those of the blind people.

Three spatial tasks were submitted to all the informants in a randomized order. The tasks have been designed with the following aims:

- eliciting spatial descriptions comparable among subjects;
- eliciting static as well as dynamic descriptions;
- displaying various situations, that might elicit different perspectives or Frames of Reference[4];
- selecting places according to different degrees of familiarity.

Therefore, we designed the following (semi-spontaneous) tasks:

*Task a.*: *bedroom description* - The subject is asked to describe his/her own bedroom as thoroughly as possible. After the subject has completed the description, the interviewer asks some common questions about the

[2] Languages that behave like English (e.g. German) are classified as *Satellite-Framed*, whereas languages like Italian (e.g. other Romance languages) are called *Verb-Framed* (Talmy, 1991).

[3] We took into account not only the educational attainment but also speakers' culture in general, assessed by the number of books they read, the movies they watch and their hobbies. The kind of place they live in (city or small town or even countryside), their 'social' life (whether they have friends and go out with them) and the degree of their autonomy (e.g. whether they need to be accompanied or supervised by someone else) have been regarded as well.

[4] For the notion of Frame of Reference, see § 4.2.2.

bedroom, e.g. *Is there any picture on the wall? Where are they? Where is the bed with respect to the door?* The answers are meant to provides us with both quantitatively and qualitatively similar data across different subjects. The bedroom is a familiar place, which is meant to elicit a (mostly) static description.

*Task b.*: *urban itinerary* - With the modality of role play, the interviewer pretends to be a tourist who meets the subject at a specific point A of his/her city and asks him/her how to reach a specific point B. The subject is also asked to take a route such that the tourist would be able to see as many touristic places as possible. After the subject has completed the description, the interviewer asks some common questions about the chosen route, e.g. about the route in Pisa: *Where is piazza dei Cavalieri with respect to piazza Santa Caterina? How far is piazza dei Miracoli from piazza Santa Caterina?* We selected two routes, one 'easier' (allegedly best known) than the other. In either cases the chosen itineraries were supposed to be known by the subjects. However, some blind subjects did not know either routes, because of their lack of movement autonomy. Therefore, in some cases the subject himself/herself proposed a short route he/she covers daily. The task proposed is meant to elicit a (mostly) dynamic description.

*Task c.*: *bird-eye city description* - The subject is asked to describe his/her city from the highest point of view possible (e.g. a bell-tower) to a tourist who has never visited it. After the subject has completed the description, the interviewer asks some common questions about the city, e.g. about Pisa: *Where is piazza del Duomo with respect to the rail station? How far are they? Is it big? What shape do you think it is?* One's own city is supposed to be a known place, but the point of view is completely unusual for both blind and sighted subjects. The task proposed should elicit a (mostly) static description and provide data to contribute to the current debate on whether blind individuals show some 'preference' for a specific spatial perspective (e.g. route over survey: see Taylor and Tversky 1992, 1996) or reference frame (Noordzij *et al.*, 2006) opposite to that chosen by sighted people.

All tasks, submitted and performed orally, were recorded. The audio files were then transcribed, using the Dragon speech recognition software with a re-speaking technique, and then manually checked. Then the transcriber, while listening again to the audio files, corrected manually all the inconsistencies and misspellings in the transcripts. Finally, the transcripts have been checked by the interviewer who actually submitted the tasks.
The transcript format used is CHAT, by the CHILDES project (MacWhinney, 2000). Although we are aware of other formats (e.g., annotation graphs; Bird and Liberman, 2001), developed especially in the area of multimodal annotations, we chose CHAT to transcribe the audio files as it is the current standard transcription system used in psycholinguistic analyses.

Figure 1 represents an example of a transcript according to the CHAT coding scheme: it is an excerpt from task a. (It. *descrizione stanza*, 'room description') performed by a blind (It. *non vedente*) male subject from Lucca (LU).

```
Non_Vedente_A.R._LU

i@Begin
@Languages: ita
@Participants: SUB Subject13nonvedente, INV Investigator
@ID: ita|descrizione stanza|SUB|36;9.25|male|||Subject13|18|
@ID: ita|descrizione stanza|INV||female|||Investigator||
@Date: 06-APR-2011
@Transcriber:Giulia
*INV: Ok, quindi descrizione della stanza propria.
Descrivi nella maniera più dettagliata possibile la tua camera da letto.
*SUB: Quindi compresi gli oggetti, la posizione degli oggetti?
*INV: Sì, tutto quello che vuoi.
*SUB: Ok, no vabbe' chiedo perché+//.
Allora intanto diciamo che la forma della stanza (…) è (eee), almeno dal punto di vista percettivo, rettangolare.
(…) è una: stanza se- (mmm), diciamo così due caratteristiche, abbastanza grande per cui viene divisa in: [x2] due.
```

Figure 1: Excerpt of a transcript in CHAT

## 4. Description of our Annotation Scheme

Bearing in mind the goals of our project (§ 1), we designed an XML annotation scheme able to capture both the specific structure of the language used by the informants (viz. Italian) and the underlying conceptual components or strategies that yield a specific spatial description (§ 3). So our approach is basically empirical, driven by actual instances of language use found in our corpus.
For the moment we decided not to adopt existing annotation languages, such as ISO-Space for instance (Pustejovsky *et al.* 2011). ISO-Space is an annotation specification, designed for capturing spatial and spatio-temporal information in natural language text. It aims at providing an inventory of how spatial information is presented in natural language such that it can be integrated by complement information coming from other modalities (e.g. GPS). ISO-Space is supposed to serve different purposes, such as, for instance: determining object location given a verbal description, constructing a route given a route description, integrating spatial descriptions with information from other media, reconstructing spatial information associated with a sequence of events, etc. (Pustejovsky *et al.* 2011: 1). We would like to briefly present two characteristics of ISO-Space, to highlight the differences between our annotation scheme and that of ISO-Space. To serve the above-mentioned purposes, ISO-Space distinguishes two major types of elements: entities (that include location and spatial entities, as well as both dynamic motion and static arrays) and spatial relations (that specify what kind of relation holds between the entities involved). As for locations, the annotation scheme of ISO-Space provides a number of elements and attributes that can be easily integrated with information deriving from other resources, such as physical feature databases and gazetteers (Pustejovsky *et al.* 2011: 3). As

for motion verbs, ISO-Space uses a classification by Muller (1998), that distinguish verbs according to their semantic meaning: *move*, *move_external* (if the motion takes place outside), *move_internal* (if the motion takes place inside), *leave*, *detach*, *deviate*, etc.
By contrast, in designing our annotation scheme, focused on the analysis of linguistic spatial categories of Italian, we adopted many distinctions commonly drawn in linguistic studies on the encoding of space, that are not included in ISO-Space (or are used differently: see, e.g., "qualitative spatial links"). Therefore, our annotation scheme is designed to analyze how language encodes space, whereas ISO-Space is designed to explicit how the space 'out-there' is captured by language.
However, with our markup language we would like to contribute to the current development of other annotation languages, such as ISO-Space, from a 'more linguistic' point of view (cf. Mani and Pustejovsky, 2012).
The XML markup language we designed encodes both morpho-syntactic and semantic categories So far 88 recordings, corresponding to task a. and b., have been annotated by three annotators.
In this section we describe the main characteristics of our markup language, using example annotations from our corpus.
Our markup language provides two major tags:

- <motion_event>;
- <localization_event>.

In other words, there is a major distinction between dynamic and static descriptions.

## 4.1 Motion event

It is every situation involving either movement or displacement[5].
In linguistics, dating back to Tesnière (1959), scholars usually distinguish between 'movement' and 'displacement'. The first term refers to the type of motion encoded by the verb (e.g. *to roll*, *slid*, *walk*, *run*); the second to the displacement, i.e. the complete shift of an entity through space. Our definition of 'motion event' embraces both distinctions. We included metaphorical motion as well: it is marked with the attribute "fictive"[6]:

<motion_event type="fictive">
*Via Duomo ci porterebbe in piazza del Giglio*[7]
</motion_event>

Every *motion_event* element usually contains other two tags:

- <motion_verb>;
- <spatial_role>.

[5] Please notice that our definition of 'motion event' differs from that by Talmy (1985), which includes both motion and stationary location.
[6] "Linguistic instances that depict motion with no physical occurrence" (Talmy 2000, I: 99).
[7] 'Via Duomo would take us in piazza del Giglio'.

### 4.1.1. Motion Verbs

By means of different attributes, we distinguish between different types of verbs. These distinctions rely mostly upon the conceptual components identified by Talmy (1985: 61): Motion: "refers to the presence *per se* in the event of motion"; Path: "is the course followed … by the Figure object with respect to the Ground object"[8]; Manner: refers to the type of motion. Therefore, the attributes of the element *motion_verb* are the following:

- dislocation: the verb lexicalizes only the Motion:

<motion_event>
<motion_verb type="disl">*Si va* </motion_verb>
*in piazza San Martino*[9]
</motion_event>

- path: the verb lexicalizes the Motion and the Path:

<motion_event>
<motion_verb type="path">*Attraversi* </motion_verb>
*il ponte*[10]
</motion_event>

- manner: the verb lexicalizes the Motion and its Manner:

<motion_event>
*Io continuo a*
<motion_verb type="manner">*camminare* </motion_verb>
*sul marciapiede*[11]
</motion_event>

- conveyance: the verb conflates the self-movement of the "carrier" entity and the "caused-movement" of the entity "carried" (e.g. to *bring*, *take*, *lead*)[12]:

<motion_event type="fictive">
*Via Duomo ci*
<motion_verb type="conv" *porterebbe* </motion_verb>
*in piazza del Giglio*[13]
</motion_event>

The last two attributes rely not on the semantics of the verbs, but on their syntactic encoding:

- construction: the verb itself does not encode the notion of Motion, which is conveyed by the pair verb-direct object instead:

<motion_event>
<motion_verb constr="y">*fare* </motion_verb>
*quindi tutta la via Grande*[14]
</motion_event>

[8] "The Figure is a *moving* or conceptually mov*able* object whose path or site is at issue; the Ground is a reference-frame, or a reference-point stationary within a reference-frame, with respect to which the Figure's path or site is characterized" (Talmy, 1983: 232).
[9] 'We go in piazza San Martino'.
[10] 'You cross the bridge'.
[11] 'I keep walking on the sidewalk'.
[12] Many scholars limit their research to intransitive verbs that encode human prototypical motion.
[13] For the translation, see footnote n. 7.
[14] Lit. '(we have) to do all via Grande', i.e. 'we have to pass through via Grande'. Besides path constructions, we found manner constructions as well: e.g. *farmi una passeggiata*, lit. 'I do (i.e. take) a walk'.

- phrasal verb: it is a phrase constituted by a verbal head and a complement represented by a ‘particle’ (originally an adverb); its syntactic cohesion is so tight that it is not possible to replace the whole phrase with only one of its parts[15]:

<motion_event type="fictive">
*dal cimitero*
<motion_verb phv="y">*andava su*</motion_verb>[16]
</motion_event>

#### 4.1.2. Spatial roles

In the literature concerning motion events, there are distinctions relative to the entities that are used as the reference-frame of the events. These distinctions have been introduced by Fillmore (1971, now 1997: 40): when talking about ‘locomotion’ (i.e. an object change of location in time), he formulates the “case-like” notions of *source*, *goal*, *path* and *location*. They do not represents conceptual elements, like Talmy’s notions (§ 4.1.1), but thematic roles: therefore they refer to the semantic function of a noun phrase with respect to its verb.

Four main distinctions pertain to spatial roles, that are marked also by the attribute indicating the part of speech (pos), such as prepositional phrase (pp), noun phrase (np), adverb (adv), etc.:

- source: is the place whence an entity departs:

<motion_event type="fictive">
<spatial_role role="source" pos="pp">*dal cimitero* </spatial_role>
<motion_verb phv="y">*andava su*</motion_verb>[17]
</motion_event>

- goal: is the destination reached by an entity:

<motion_event type="fictive">
*Via Duomo ci*
<motion_verb type="conv"> *porterebbe* <motion_verb>
<spatial_role role="goal" pos="pp">*in piazza del Giglio*</spatial_role>[18]
</motion_event>

- path: is the course followed by an entity during its motion (cf. Talmy’s definition, § 4.1.1):

<motion_event>
<motion_verb type="path">*Attraversi* </motion_verb>
<spatial_role role="path" pos="np">*il ponte*</spatial_role>[19]
</motion_event>

- manner: it is the manner of motion[20]:

<motion_event>
<motion_verb type="path">*Sono partito* </motion_verb>
<spatial_role role="manner" pos="adv">*di scatto* </spatial_role>[21]
</motion_event>

[15] The definition is adapted from Simone (1996: 49), where it is applied to the Italian ‘verbi sintagmatici’ (‘syntagmatic verbs’).
[16] ‘From the graveyard it [*scil.* the road] went uphill’.
[17] For the translation, see footnote n. 16.
[18] For the translation, see footnote n. 7.
[19] For the translation, see footnote n. 10.
[20] We introduced this new spatial role because in the Verb-framed languages (such as Italian), the Manner is usually encoded - when encoded at all - by an adjunct (§ 2).
[21] Lit. ‘I left suddenly’, i.e. ‘I bolted’.

### 4.2 Localization event

It is every situation involving a stationary location of an entity (Figure) with respect to other entities (Ground).[22]

Every *localization_event* element usually contains other two elements:

- <localization_verb>;
- <spatial_relation>.

#### 4.2.1. Localization Verbs

Many languages have a series of verbs that describe the static position of a Figure with respect to a Ground entity (cf. Ameka and Levinson, 2007). See, for instance, the following postural verbs: En. *to lie, sit, stand, hang*; Ger. *liegen* ‘to lie’, *sitzen* ‘to sit’, *stehen* ‘to stand’, and their respective dynamic counterpart *legen* ‘to lay’, *setzen* ‘to sit’, *stellen* ‘to stand’ (Rüsch, 2010).

In Italian there are very few verbs of position and they are rarely used: e.g. *giacere* ‘to lie’ (which pertains to the literary register), *sedere* ‘to sit’. In every day communications, periphrastic constructions are preferred: they are formed by the verbs *stare* ‘to stay’ or *essere* ‘to be’, and a former past participle or an adverbial: *stare sdraiato* ‘to lie’, *stare seduto* ‘to sit’, *stare in piedi* ‘to stand’, *essere/stare appeso* ‘to hang’. The spatial information conveyed by these verbs and constructions in Italian is optional; moreover it prototypically pertains to human beings.

Since the semantics of Italian localization verbs does not entail many distinctions, at the moment our markup language does not provide further attributes for localization verbs[23]. In fact, in our corpus we found mostly the verb *essere/esserci* ‘to be/there be’ or verbs like *avere* ‘to have’, and *trovare/trovarsi* ‘to find/be’:

<localization_event>
*Su questo mobiletto*
<localization_verb>*c’è* </localization_verb>
*un cestino*
</localization_event>[24]

#### 4.2.2. Spatial Relations

In the Western tradition of spatial studies, a lot of attention has been paid to (static) spatial relations, at least since the pioneer work by Piaget and Inhelder (1948) on the development of spatial representations in human beings. They showed that the first spatial concepts acquired are the topological notions of proximity, order, closure, and continuity; only much later, children understand the Euclidean notions of metric distance and

[22] Linguistic description of space, both static and dynamic, is highly relational (Meini, 2009; Mani and Pustejovsky, 2012): usually we locate an entity x (Figure) by reference to the location of entity y (Ground). Similarly, we describe an entity w (Figure) as moving in relation to an entity z (Ground); a dynamic description is not relational when it involves only the ‘movement’ of the Figure (§ 4.1).
[23] However, distinctions concerning localization verbs (even postural verbs for the annotation of languages different from Italian) could be easily added into our markup language.
[24] ‘On this little table there is a bin’.

angle, and, at last, they are able to grasp geometrical projective relations.

Irrespective to the different and opposing theoretical paradigms developed on the basis of Piaget and Inhelder's findings (§ 1), the distinction between topological and projective relation is still maintained in linguistic researches[25]. Therefore, in our markup language, we distinguish these two main types of spatial relations[26]. Each entity involved in both topological and projective relations are marked by the attribute indicating the part of speech (§ 4.1.1).

Following studies on spatial relations (Becker, 1997; Meini, 2009) we distinguish six main kinds of TOPOLOGICAL relations[27]:

- <u>at place</u>: the Figure is localized with respect to the Ground entity without any further spatial information:

<spatial_relation type="top" subtype="at place" pos="pp">
*alla finestra*[28]
</spatial_relation>

- <u>inner</u>: the Figure is localized with respect to the inner subspace of the Ground entity:

<spatial_relation type="top" subtype="inner" pos="pp">
*nella mia stanza da letto*[29]
</spatial_relation>

- <u>neighbouring</u>: the Figure is localized with respect to the subspace surrounding the Ground entity:

<spatial_relation type="top" subtype="neigh" pos="pp">
*vicino al letto*[30]
</spatial_relation>

- <u>boundary</u>: the Figure is localized with respect to the boundary of the Ground entity:

<spatial_relation type="top" subtype="boundary" pos="pp">
*Su questo mobiletto*[31]
<spatial_relation>

- <u>exterior</u>: the Figure is localized with respect to the exterior subspace of the Ground entity:

<spatial_relation type="top" subtype="ext" pos="pp">
*esternamente alla portafinestra*[32]
</spatial_relation>

- <u>between</u>: the Figure is localized with respect to a 'complex' Ground composed of disjunct entities:

<spatial_relation type="top" subtype="betw" pos="pp">
*tra il letto e la finestra*
</spatial_relation>[33]

PROJECTIVE relations suppose the notion of direction and a system of axis. Therefore, we distinguish three main kinds of projective relations, according to the axis involved:

- lateral: <spatial_relation type="proj" axis="lat" pos="pp">
  *sulla destra*[34]
  </spatial_relation>
- sagittal: <spatial_relation type="proj" axis="sag" pos="pp">
  *di fronte al campo da tennis*[35]
  </spatial_relation>
- vertical: <spatial_relation type="proj" axis="ver" pos="pp">
  *sopra il letto*[36]
  </spatial_relation>

For every projective relation we also indicate the frame of reference (henceforth FoR)[37].

The notion of FoR has been introduced in linguistic analyses by Levinson (2003: 24), who provides the following definition, quoting from Irvin Rock: "a unit or organization of units that collectively serve to identify a coordinate system with respect to which certain properties of objects, including the phenomenal self, are gauged"[38].

Three main FoR have emerged from Levinson (2003: 38 ff.) analysis (the corresponding attribute of our annotation is in brackets):

- Intrinsic (FoR="intr"): the coordinates are determined by the so-called 'inherent features' (i.e. sidedness or facets) of the Ground entity;
- Relative (FoR="rel"): directions are assigned to Figure and Ground by the coordinates fixed on a distinct 'viewpoint';
- Absolute (FoR="absol"): this FoR is based on fixed bearings, such as cardinal points.

Sometimes linguistic data are not enough to tell an intrinsic from a relative FoR, since in many languages there are not linguistic items specific for the two domains. In these cases, the annotator cannot only rely on the linguistic text, but has to supplement it with extra-textual information.

Since the kind/s of FoR usually employed in a language is/are culturally determined (Levinson, 2003), the annotation will allow us to analyze which FoR is mostly used (or mostly avoided) by Italian speakers. We will also evaluate whether the choice of a specific FoR is influenced by the kind of spatial description, e.g. a small

[25] For more recent studies on the development of prelinguistic spatial concepts (and its relationship with the acquisition of spatial language), see e.g. Mandler (2004).

[26] For a discussion of how Italian, English and Spanish encode both topological and projective relations by means of prepositions or adverbials, see Meini (2009).

[27] For reasons of space, in this section we quote only the annotation relative to the spatial relation and not that of the whole localization event. In the translation, the full sentence is provided.

[28] '(There are curtains) at the window'.

[29] 'In my bedroom (there is a door)'.

[30] '(it [*scil.* the armchair] is) near the bed'.

[31] For the translation, see footnote n. 24.

[32] 'Outside the French doors (there is also a terrace)'.

[33] 'Between the bed and the window (there is an armchair)'.

[34] '(There is a television) on the right'.

[35] 'Opposite to the tennis court (there is shop)'.

[36] '(This fan is) above the bed'.

[37] For obvious reasons, the specification of the *axis* is not needed in case of an absolute FoR.

[38] It follows that FoR must not be confused with the (kind of the) origin of the coordinate system: e.g. the opposition 'egocentric' vs 'allocentric'.

room vs wide spaces (such as a city). Finally, the analysis of the FoR in our corpus will reveal whether there are any differences between blind and sighted individuals, hence whether the different experiential base influences somehow the choice of the FoR.
As for our annotation, we drew a sketch of the speakers' bedrooms (task a., Figure 2), and we checked the route followed by the informants with Google Map or Google Earth (task b., Figure 3); we suppose that Google Earth will be an useful tool when annotating the descriptions of task c. as well.

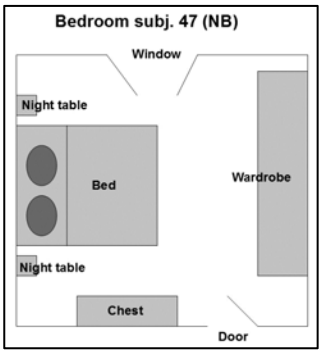


Figure 2: Sketch of subj. 47's (Not Blind) bedroom

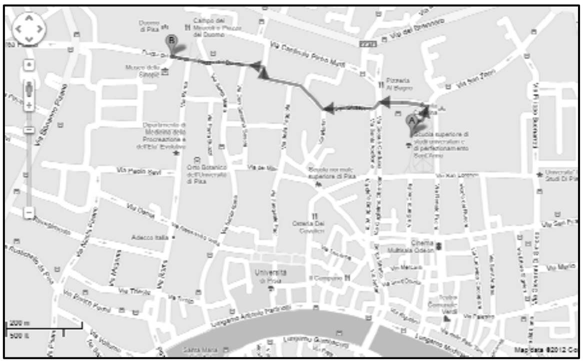

Figure 3: Route followed by subj. 23 (Google Map - Pisa)

### 4.3 Spatial Attributes

Finally, the markup language provides the element *attribute*, which concerns mostly spatial attributes, such as: shape (e.g. *ottagonale* 'octagonal'), size (e.g. *piccolo* 'small'), material (e.g. *di legno* 'wooden'). Moreover, we included the attribute *metric distance*, to verify whether the speakers' more or less accuracy in distance comparison (task b. and c.) could be related to the different experiential base of blind and sighted individuals (cf. also Noordzij *et al.* 2006).
As with spatial roles and spatial relations, the attribute elements are marked by the indication of the part of speech (§ 4.1.1, § 4.2.2).

## 5. Conclusion

In this paper we reported on a new annotated resource currently being developed to analyze spatial information in a corpus of spoken Italian. The resource consists of various material: spatial descriptions made by Italian speakers (and the corresponding transcripts in CHAT format) are associated with sketches of rooms and maps of routes.
The spatial annotation scheme and markup language we designed aims at providing a comprehensive tool that allows the researcher to highlight:

- effects of the specific lexical and grammatical features of the language spoken by the informants (in this case Italian) on the encoding of space;
- differences in the encoding of space related to:
  - speakers' sociolinguistic variables, such as age, gender, dialect spoken, culture, etc. (§ 2);
  - different experiential base between blind and sighted individuals (§ 1).

In designing the markup language, we referred mainly to the linguistic literature on space and to psycholinguistic studies on the language of the blind. However, our approach is empirical: it means that among the overall amount of distinctions drawn in linguistic literature, we chose those relevant to the actual instances of language use found in our corpus.
In the very next future we are going to:

- enrich our markup language with the annotation of nouns, functioning especially as Ground (Herskovits, 1986; Vandeloise, 1991; Frank, 1997; Meini, 2009; Bateman *et al.*, 2010);
- enrich the semantic annotation of verbs, especially the motion verbs (e.g. Italian path verbs encode many different 'experiential' situations);
- carry out inter-coder agreement tests;
- automatically PoS-tag and lemmatize the corpus;
- explore possible synergies between our annotation scheme and ISO-Space;
- align speech recordings with transcripts and other coding layers using annotation graphs.

Our linguistic analyses aims at providing more evidence about spatial language use in Italian.
All the resources developed by the project will be publicly and freely available, and they should be of interest to a wide scientific community. The resources will be released with Creative Commons Attribution-Share Alike license (see http://creativecommons.org/).

## 6. References

Ameka, F. and Levinson, S.C. (2007). Introduction. The Typology and Semantics of Locative Predicates: Posturals, Positionals, and Other Beasts. *Linguistics*, 45: pp. 847--871.

Bateman, J.A., Hois, J., Ross, R., Tenbrink, T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14): pp. 1027--1071.

Becker, A. (1997). Theoretical framework. In Becker, A. and Carroll, M. (Eds.), *The acquisition of spatial relations in a second language*. Amsterdam-Philadelphia: J. Benjamins, pp. 13--33.

Bird, S. and Liberman, M. (2001). A Formal Framework for Linguistic Annotation. *Speech Communication* 33(1,2): pp. 23--60.

Fillmore, C. (1971). *Santa Cruz lectures on deixis*. Bloomington: Indiana University Linguistics Club [now Fillmore, C. (1997). *Lectures on Deixis*. Stanford: CSLI Publications].

Frank, A.U. (1997). Spatial ontology: A geographical information point of view. In: Stock, O. (Ed.), *Spatial and Temporal Reasoning*. Dordrecht: Kluwer Academic Publishers, pp. 135--153.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: CUP.

Jackendoff, R. (2002). *Foundations of Language*. Cambridge, MA: MIT Press.

Landau, B., Dessalegn, B. and Goldberg, A.M. (2010). Language and Space: Momentary Interactions. In Evans, V. and Chilton, P. (Eds.), *Language, Cognition and Space: The State of the Art and New Directions*. London: Equinox Publishing, pp. 51--77.

Levinson, S.C. (2003). *Space in language and cognition: explorations in cognitive diversity*. Cambridge: CUP.

Levinson, S. and Wilkins, G. (Eds.) (2006) *Grammars of Space*. Cambridge: CUP.

MacWhinney, B. (2000), *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandler, J. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford: OUP.

Mani, I. and Pustejovsky, J. (2012), *Interpreting Motion: Grounded Representations for Spatial Language*. Oxford: OUP.

Marotta, G. (2010). Introduction. In Marotta, G., Lenci, A., Meini, L. and Rovai, F. (Eds.), *Space in Language 2009. Proceedings of the Pisa International Conference*. Pisa: ETS, pp. 11--25.

Meini, L. (2009). *Dimensioni dello Spazio nelle Preposizioni. Un'Analisi Empirica sull'Italiano L2*. Pisa: PLUS.

Miller, G.A. and Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge: CUP.

Muller, P. (1998). A qualitative theory of motion based on spatio-temporal primitives. In Cohn, A.G., Schubert L., and Shapiro, S.C. (Eds.), *KR'98: Principles of Knowledge Representation and Reasoning*. San Francisco, CA: Morgan Kaufmann, pp. 131--141.

Nordzij, M. Zuidhoek, S. and Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, 100, pp. 321--342.

Piaget, J. and Inhelder, B. (1948). *La représentation de l'espace chez l'enfant*, Paris: Presses Universitaires de France.

Pustejovsky, J., Moszkowicz, J. and Verhagen, M. (2011). ISO-Space: The Annotation of Spatial Information in Language. In Bunt, H.C. (Ed.), *Proceedings of the 6th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6), January 11–12, 2011, Oxford, UK*. Oxford: University of Oxford, pp. 1--9.

Rüsch, N. (2010). Placement events in German as a second language: *Legen*, *stellen*, and *setzen* and their static counterparts. In Marotta, G., Lenci, A., Meini, L. and Rovai, F. (Eds.), *Space in Language 2009. Proceedings of the Pisa International Conference*. Pisa: ETS, pp. 629--641.

Simone, R. (1996). Esistono verbi sintagmatici in italiano?. *Cuadernos de Filología Italiana*, 3, pp. 47--61.

Slobin, D.I. (1985). Crosslinguistic evidence for the language-making capacity. In Slobin, D.I. (Ed.), *The crosslinguistic study of language acquisition*, *vol. II*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 1157--1256.

Slobin, D.I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In Strömqvist, S. and Verhoeven, L. (Eds.), *Relating Events in Narrative*, *vol. II: Typological and Contextual Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 219--257.

Talmy, L. (1983). How language structures space. In Pick, H.L. and Acredolo, L.P. (Eds.), *Spatial orientation: theory, research, and application*. New York-London: Plenum Press, pp. 225--282.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In Shopen, T. (Ed.), *Language Typology and Syntactic Description*. Cambridge: CUP: pp. 57--149.

Talmy, L. (1991). Path to realization: A typology of event conflation. In Sutton, L.A, Johnson, C. and Shields, R.B. (Eds.), *Proceedings of the seventeenth annual meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, pp. 480--519.

Talmy, L. (2000). *Toward a Cognitive Semantics*, 2 *Vols.*. Cambridge, MA: MIT Press.

Taylor, H.A. and Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory & Language*, 31, pp. 261--292.

Taylor, H.A., and Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory & Language*, 35, pp. 371--391.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.

Tyler, A. and Evans, V. (2003). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: CUP.

Vandeloise, C. (1991). *Spatial Prepositions: A Case Study from French*. Chicago: The University of Chicago Press.

# Empirically Validating VerbNet Using SemLink

**Claire Bonial, Weston Feely, Jena D. Hwang and Martha Palmer**

Department of Linguistics, University of Colorado at Boulder
Hellems 290, 295 UCB, Boulder CO 80309-0295
{claire.bonial, weston.feely, hwangd, martha.palmer}@colorado.edu

**Abstract**

This research describes efforts to empirically validate a lexical resource, VerbNet, using the PropBank annotations found in the SemLink corpus. As a test case, we examine the frequency with which verbs in SemLink appear in the Caused-Motion syntactic frame: NP-V-NP-PP (e.g., *She poured water into the bowl*). To do this, we find the frequency with which a given verb is used in this construction, we then determine each verb's VerbNet class membership, and compare the overall frequency of the Caused-Motion construction in the verb class to how the verbs' behavior is currently represented in VerbNet. We find evidence that VerbNet's current classification fails to capture generalizations about the likelihood of a class' compatibility with the Caused-Motion construction. Specifically, classes where Caused-Motion is currently represented in VerbNet as a characteristic syntactic frame were found to have a lower frequency of realization in that frame than other classes where Caused-Motion is not represented. We therefore suggest augmenting VerbNet's classification with information on the probability that a class will participate in a certain syntactic frame, and given the challenges of this research, offer potential improvements for increasing the interoperability of VerbNet.

## 1. Introduction

VerbNet (VN) (Kipper et al., 2008) is a classification of English verbs, expanded from Levin's (1993) classification. VN serves as a valuable lexical resource, facilitating a variety of Natural Language Processing (NLP) tasks such as semantic role labeling (Swier and Stevenson, 2004), inferencing (Zaenen et al., 2008), and automatic verb classification (Joanis et al., 2008). However, most classes have yet to be validated by research examining the actual usage of verbs. In this research, we take the first steps towards empirically validating VN by comparing instances of the Caused-Motion construction (Goldberg, 1995) in the SemLink corpus (Palmer, 2009) to its representation in VN. We selected the Caused-Motion construction as the focus of this research in order to expand upon previous research on this construction (Bonial et al., 2011c).

### 1.1. VerbNet Background

Class membership in VN is based on a verb's compatibility with certain syntactic frames and alternations. For example, all of the verbs in the *Spray* class have the ability to alternate the Theme or Destination as a noun phrase (NP) object or as a prepositional phrase (PP): *Jessica loaded the boxes into the wagon*, or *Jessica loaded the wagon with boxes*.

VN's structure is somewhat hierarchical, comprised of superordinate and subordinate levels within each verb class. In the top level of each class, syntactic frames that are compatible with all verbs in the class are listed. In the lower levels, or 'sub-classes,' additional syntactic frames may be listed that are restricted to a limited number of members. In each class and sub-class, an effort is made to list all syntactic frames in which the verbs of that class can be grammatically realized. Each syntactic frame is detailed with the expected syntactic phrase type of each argument, thematic roles of arguments, and a semantic representation; for example:

**Frame** NP V NP PP.destination
**Example** Jessica loaded boxes into the wagon.
**Syntax** Agent V Theme Destination
**Semantics** Motion(during(E), Theme)
Not(Prep-*into*(start(E), Theme, Destination))
Prep-*into*(end(E), Theme, Destination)
Cause(Agent, E)

### 1.2. SemLink and PropBank Background

The SemLink corpus (Palmer, 2009; Loper et al., 2007) consists of 112,917 instances of the Wall Street Journal, each annotated with its corresponding VN class. Each instance is further annotated with PropBank (Palmer et al., 2005) arguments, which are numbered arguments that correspond to verb-specific roles. For example, these are the potential roles to be assigned for the verb *load*:

**Roleset id**: load.01, *cause to be burdened*, VN class: 9.7-2
**Roles**:
Arg0: loader, agent (VN role: 9.7-2-agent)
Arg1: beast of burden (VN role: 9.7-2-destination)
Arg2: cargo (VN role: 9.7-2-theme)
Arg3: instrument

Note that each verb sense, or 'Roleset' is mapped to its corresponding VN class, and each of the PropBank roles are mapped to VN thematic roles where possible. This roleset also demonstrates a sort of mismatch between PropBank and VN's treatment of *load*: PropBank treats the instrument as a numbered argument, whereas VN doesn't list an instrument as a semantic role for this verb. Within the SemLink corpus, these mappings are made explicit such that with each instance, both PropBank and VN thematic roles are given for each argument. SemLink also contains mappings between PropBank rolesets, VN classes and FrameNet (Fillmore et al., 2002) frames, as well as corresponding mappings between PropBank arguments, VN thematic roles and FrameNet frame elements. Thus, SemLink is a resource created with the intent of allowing for interoperability amongst these resources, yet some difficulties remain in taking full advantage of this resource. We will discuss a few of the challenges we faced in Section 2.3.

# 2. Empirically Validating VerbNet

## 2.1. The Caused-Motion Test Case

The PropBank annotations and mappings to VN found in SemLink allow us to examine whether or not the characterizations of syntactic behavior found in VN are truly representative of a verb's behavior in actual usage. To take the first step in the process of empirically validating VN, we chose to examine the Caused-Motion (CM) construction. We selected this construction in order to expand upon previous research (Bonial et al., 2011c), in which we compared VN's representation of CM firstly to manual annotations of CM in a portion of the Wall Street Journal, and secondly to automatic detections of CM using Hierarchical Bayesian Modeling. The starting point of this research was a manual examination of VN and a thorough annotation of the status of each class with respect to the CM construction, performed by two linguistic graduate students and adjudicated by a third. This effort revealed a number of shortcomings in VN and motivated hypotheses about the expected patterns of use of the CM construction across the classes. We continue to make use of these annotations in our current research. A complete description of the annotation process is given in the next section.

### 2.1.1. Annotation Guidelines and Results

The first goal of our manual annotation of VN classes was to determine which classes currently represent CM in one of their frames. To this end, we identified which classes contain the following frame:

NP[Agent/Cause]-V-NP[Patient/Theme]-
PP[Source/Destination/Recipient/Location]

These frames correspond to classes such as *Slide*, with its frame NP-V-NP-PP.Destination: *Carla slid the books to the floor*. We also examined classes with the patterns NP-V-NP-PP.Oblique, NP-V-NP-PP.Theme2, and NP-V-NP-PP.Patient2. In these classes, annotators had to judge whether the final PP was indicative of CM. For example, the *Breathe* class contains the frame NP-V-NP.Theme-PP.Oblique, *The dragon breathed fire on Mary*, which is an example of CM, whereas the same basic frame in the *Other_cos* class is not: NP V NP PP.Oblique, *The summer sun tanned her skin to a golden bronze.*

In addition to identifying which classes contain a CM frame, we also annotated which classes were potentially compatible with CM for either all verbs in the class or only some verbs. The 'some' classification has the drawback that it may be applied to classes with very different proportions of compatible verbs. A secondary determination was whether or not the class was compatible with CM as part of its core semantics, or if it was compatible with CM because it was coercible into the construction. A verb was considered 'compatible with CM' and 'not coerced' if the verb could be used in the CM construction grammatically and its semantics, as reflected in VN's semantic predicates, involved a CAUSE predicate in combination with another predicate such as CONTACT, TRANSFER, (EN)FORCE, EMIT, TAKE_IN (predicates potentially involving movement along some path). For example, although CM is not already included as a frame for the *Bend* class containing the verb *fold*, the semantics of this class include CAUSE and CONTACT, and the verb can be used in a CM construction: *She folded the note into her journal.* Therefore, this class would have been considered 'compatible with CM' but 'not coerced.' Conversely, a verb was considered 'compatible with CM' and 'coerced' if the verb could be used in the CM construction, yet its semantics, again as reflected in VerbNet, did not involve CAUSE and MOVEMENT ALONG A PATH (e.g., the verb *wiggle* of the *Body_internal_motion* class: *She wiggled her foot out of the boot*).

In summary, as presented in Table 1, we annotated each class according to whether (1) the CM construction was already represented in VN for this class, (2) the construction was possible for all, some, or none of the verbs in that class, and (3) the verbs of any class that was compatible with CM were coerced into the construction or not. The classification for (3) was made regardless of whether 'all' verbs or only 'some' were compatible with CM. This determination was made uniformly for a class; that is, there were no classes in which some CM-compatible verbs were considered coerced and some were not.

Notably, we identified 206 classes where at least some of the verbs in that class are compatible with the CM construction; however, VN currently only recognizes the CM construction in 58 classes. There were several classes of interest: first, in Group 3, although it may seem unusual that CM is represented in 6 classes of VN where we found that only 'some' verbs were compatible with CM (e.g., *Cheat* class) these were cases where only more restricted subclasses are compatible with CM, and this syntactic frame is listed for that subclass. This suggests that VN's subclasses may provide a more precise characterization of which verbs are compatible with a construction. Secondly, in Group 4, we identified 18 classes in which all verbs were compatible with CM without coercion; thus, these classes could likely be improved by the addition of the CM syntactic frame. Additionally, in Group 5, we found 30 classes in which all verbs are coercible into the CM construction; however, the actual likelihood of a verb in those classes occurring in a CM construction remains to be investigated in the following sections. Like those classes where it was determined that only 'some' verbs are compatible with CM, usefully incorporating the CM construction into classes that require coercion relies on accurately determining the probability that verbs in those classes will actually appear in the CM construction.

For those classes in which 'all' verbs are compatible with CM, our intuition was that some aspect of the verb's semantics either inherently includes CM or allows the verb to be coerced into the CM construction. Conversely, for those classes in which no verbs are compatible with CM, presumably some aspect of the verb's semantics is logically incompatible with CM. Although pinpointing precisely what

| Grouping | VN class example | # of classes like this | CM rep. in VN | CM is possible for... | CM is coerced |
|---|---|---|---|---|---|
| 1 | *Banish* | 50 | Yes | All | No |
| 2 | *Nonverbal_Expression* | 2 | Yes | All | Yes |
| 3 | *Cheat* | 6 | Yes | Some | No |
| 4 | *Exhale* | 18 | No | All | No |
| 5 | *Hiccup* | 30 | No | All | No |
| 6 | *Fill* | 46 | No | Some | No |
| 7 | *Wish* | 54 | No | Some | Yes |
| 8 | *Matter* | 64 | No | None | N/A |

Table 1: Annotation results–VN classes segmented on the basis of whether or not CM is already represented in VN, CM is possible for 'all/some/none' of the verbs in a class, and for those verbs that are compatible with CM, whether they are compatible only through coercion or not.

aspect of a verb's semantics makes it either compatible or incompatible with CM may not be possible, we can investigate whether or not our intuitions are supported by examining the actual frequencies of CM constructions for given verbs or a given class.

Given the information obtained in these annotations, we had certain hypotheses about what the distribution of CM would be in the SemLink corpus. First, Group 4 in Table 1 is of special interest: these 18 classes were found to be compatible with CM for all members, and this compatibility was thought to be part of the verbs' core behavior, yet CM is not represented as part of their core behavior currently in VN. In our investigation of CM in SemLink, we hypothesize that verbs in these classes will have a comparatively high frequency of CM. If this is the case, it will be especially important to add CM to VN's characterization of verbal behavior in these classes. Secondly, we hypothesize that the CM frequency will be highest in classes where 'All' verbs are compatible, second-highest in classes where 'Some' verbs are compatible, and lowest in classes where 'None' of the verbs are compatible with CM. Similarly, we hypothesize that the CM frequency will be highest in classes where verbs are compatible with CM as part of their core behavior, as opposed to classes where verbs must be coerced into the construction. To investigate these more general hypotheses, we regrouped the 8 groupings above into more coarse-grained segments based on whether 'all/some/none' of the verbs in the class are compatible, and according to whether that compatibility is 'core' or 'coerced'. If the verb was simply incompatible with CM, it also fell into the 'none' category, where no verbs in a class were compatible with CM as either part of their core or coerced behavior. The following table summarizes the results of this secondary partitioning.

| Grouping | Class Example | # of classes |
|---|---|---|
| **All Allowed** | *Bring* | 106 |
| **Some Allowed** | *Lodge* | 100 |
| **None Allowed** | *Try* | 64 |
| **Not Coerced** | *Put* | 120 |
| **Coerced** | *Wink* | 86 |

Table 2: Regrouping according to whether 'all/some/none' of the verbs are compatible and type of compatibility ('coerced' or 'not coerced') with CM

### 2.2. Method: Gathering Data from SemLink

To obtain the token frequency with which a certain verb is realized in the CM construction, we searched a segment of 81,585 SemLink instances for particular patterns that correspond to CM.[1] We began with the desired pattern of syntactic constituents, searching for constructions of the basic type: NP-V-NP-PP,[2] using the Penn Treebank (Marcus et al., 1993) syntactic information. We then narrowed these results according to the desired semantic roles, ideally searching for the pattern: NP.agent/cause V NP.theme PP.destination/source/direction.

However, we could not limit our search to where this pattern was present in SemLink, as this would preclude instances where CM was not already recognized in VN, thereby undermining our purpose. For example, in the following SemLink instance, *staple* is akin to *put* in meaning and usage: *She staples polaroid snapshots to the outside of each hatbox*; thus, it should be counted as an instance of the CM construction. The verb *staple* is a member of the *Tape* class in VN, which does not explicitly recognize CM in its thematic roles (in the sense that it does not use roles that we would think of as prototypical to CM: Agent, Theme, Destination/Source). According to the VN thematic roles, this instance is instead characterized by the pattern NP.Agent V NP.Patient1 PP.Patient2.

For this reason, we turned to the PropBank annotations to narrow our original results to those instances that had particular kinds of final prepositional phrases that are indica-

[1]81,021 of these instances had mappings to current VN classes, others were mapped to outdated class numbers or the class numbers contained errors; therefore, these 81,000 were the focus of our analysis.

[2]The noun tags used were more nuanced than this: we allowed for the NP positions to consist of the Penn Treebank tags NP, NN, NNS, or PRP. We allowed the subject arguments to vary even more extensively, as we found that subjects could also be S-node and other constituent types; thus, we focused primarily on the post-verbal information.

tive of CM. PropBank's numbered arguments range from 0-5, where Arg0 and Arg1 correspond to Dowty's (1991) prototypical agent and patient respectively. Args 2-5 are verb specific; thus, there is no particular correspondence between the argument number and a single thematic role. In the SemLink annotations, to disambiguate different types of numbered arguments, Args 2-5 are optionally, but not always, accompanied by the most fitting modifier (ArgM) label. Where these arguments are prepositional phrases, they are also labeled with the preposition; for example:

1. She staples polaroid snapshots to the outside of each hatbox.
   **SemLink Annotation**:
   $\text{She}_{\text{ARG0[Agent]}}$ $\text{staples}_{\text{RELATION}}$
   $\text{polaroid\_snapshots}_{\text{ARG1[Patient1]}}$
   $\text{to\_the\_outside\_of\_each\_hatbox}_{\text{ARG2[Patient2]}-\text{to}}$

For certain instances, there was also the possibility that neither VN nor PropBank recognized the CM construction in its thematic roles or numbered arguments. In such cases of CM, the final preposition would be annotated with an ArgM, such as Direction or Location, instead of a numbered argument.[3] We used both the modifier tags where available and prepositions accompanying Args 2-5 in order to narrow our results to likely CM constructions. Specifically, we excluded instances with prepositional phrases that had modifier tags of a type that we felt could not be part of a CM construction, and we excluded prepositional phrases that had prepositions incompatible with CM. The following table illustrates the overall method of our search.

| Syntactic Frame | PP Args Included | PP Modifier Exclusion | PP Type Exclusion |
|---|---|---|---|
| V NP PP | Arg# | Purpose | by |
| | ArgM-DIR | Extent | for |
| | ArgM-LOC | Manner | with |
| | | Temporal | as |
| | | Reciprocal | per |

Table 3: Data gathering process

Once we had gathered likely CM instances accordingly, we examined the frequency of the construction for a given verb, verb class, and CM grouping type given in the manual annotations discussed in the previous section. Results are discussed in Section 3.

### 2.3. Interoperability Challenges

Despite ongoing efforts to map PropBank and VN in a complementary fashion that would allow us to empirically validate VN, we faced a variety of challenges in our investigation of CM's representation in VN. First, while VN contains valuable syntactic frames for each verb class, these syntactic frames are often limited to prototypical syntactic constituents, which makes it difficult to match SemLink instances with a particular syntactic frame in VN. For example, in SemLink, a verb's subject could be realized as a clause where VN expects only an NP. Similarly, relativizers are often dropped in SemLink instances where the VN frame specifies the lexical item 'that' in the position of the relativizer. Investigating the syntactic representation of CM in VN would have been far easier if a more complete, empirically motivated set of syntactic frames were available for each VN class, or if the syntactic frames were simply more general, allowing for different types of subjects or relativizers to fill certain syntactic slots.

Fortunately, concurrent work with this project includes an experiment to enumerate all the syntactic frames in SemLink, grouped by their VN class assignments. This statistical information on syntactic frames in SemLink could be added to VN, not only to expand the number of possible syntactic frames in each VN class with empirically observed additions, but also to provide a SemLink frequency for each frame.

The 'mismatches' between VN and PropBank, mentioned in Section 1.2, were also challenging: VN assigns thematic roles to core arguments of the verb, but these thematic role assignments don't always have a corresponding numbered PropBank argument, so these are instead assigned an ArgM label. Conversely, VN sometimes does not have a thematic role assignment where PropBank assigns a numbered argument. These discrepancies in semantic role labels are a challenge to empirically validating VN using SemLink, and demonstrate that such discrepancies will always exist where two different resources have even slightly varying views of language. Given this inescapable difficulty, we suggest ways in which to facilitate interoperability with other resources in our final section on future work.

## 3. Results

Table 4 gives the results organized by the numbered groupings determined by the manual CM annotations of VN. The Total Frames reported is the total number of occurrences of a particular group of verbs, the CM Frequency is simply the number of CM occurrences divided by the number of total occurrences. Again, these groups are organized according to whether (1) CM is already represented in VN for that verb's class, (2) all, some, or none of the verbs in that class are compatible with CM and (3) the verbs in that class are compatible with CM as part of their core behavior, or only through coercion. The overall frequency of CM in SemLink is 0.088 (7204 CM frames detected in all, out of a total of 81585 instances), so classes where the frequency is higher than this are inclined to CM.

If VN's representation were already perfectly adequate, we would expect Group 1, where CM is already represented as a characteristic frame of the VN class and human annotators agreed that all verbs were compatible with CM as part of their core behavior, to have the highest frequency of CM usages. However, it is Group 4 that is characterized by the highest frequency of CM usages, despite the fact that the overall SemLink frequency of Group 4 verbs is lower than that of Group 1. Thus, our first hypothesis that Group 4 would have a comparatively high frequency of CM usages is supported. This reflects a gap in VN's representations, as

[3] At the time of SemLink's annotation, PropBank modifier arguments did not include a separate tag for a role that would be characterized as a destination or goal.

| Grouping | CM counts | Total Frames | CM Frequency |
|---|---|---|---|
| 1 - CM rep. in VN/ All compatible with CM/ Core | 2898 | 30288 | 0.0957 |
| 2 - CM rep. in VN/ All compatible with CM/ Coerced | 2 | 61 | 0.0328 |
| 3 - CM rep. in VN/ Some compatible with CM/ Core | 170 | 2254 | 0.0754 |
| 4 - CM not rep. in VN/ All compatible with CM/ Core | 526 | 3371 | 0.1560 |
| 5 - CM not rep. in VN/ All compatible with CM/ Coerced | 251 | 1678 | 0.1496 |
| 6 - CM not rep. in VN/ Some compatible with CM/ Core | 1396 | 12483 | 0.1118 |
| 7 - CM not rep. in VN/ Some compatible with CM/ Coerced | 1005 | 18169 | 0.0553 |
| 8 - CM not rep. in VN/ None compatible with CM/Incompatible | 777 | 12717 | 0.0611 |

Table 4: Frequencies of CM, grouped by CM manual annotation types: whether or not CM is currently represented in VN or not, whether all/some/none of the verbs in a class are compatible with CM, and whether that compatibility is core behavior or coerced.

human annotators also considered all verbs in these classes to be compatible with CM as part of their core behavior. Examples of usages that fall into this group would be the *staple* example discussed above (Example 1), and other illustrative examples are provided below:

2. A court in Jerusalem sentenced a Palestinian to 16 life terms for *null* forcing a bus off a cliff July 6, killing 16 people, Israeli radio reported.
   **SemLink Annotation**: *null*$_{\mathrm{ARG0[Agent]}}$ forcing$_{\mathrm{RELATION}}$ a_bus$_{\mathrm{ARG1[Patient]}}$ off_a_cliff$_{\mathrm{ARG2[Proposition]-off}}$

3. Turner Broadcasting System Inc. said it formed a unit to make and distribute movies to theaters overseas.
   **SemLink Annotation**: A_unit$_{\mathrm{ARG0[Agent]}}$ distribute$_{\mathrm{RELATION}}$ movies$_{\mathrm{ARG1[Theme]}}$ to_theaters_overseas$_{\mathrm{ARG2[Recipient]-to}}$

Although these seem to be fairly clear instances of CM, the behavior of these verbs is simply interpreted differently in VN; other aspects of their semantics are highlighted instead of CM. However, ideally, VN should represent verb behavior in a way that is informative about which usages are dominant and which are grammatical but rare. Currently, VN does not make any distinction between syntactic frames based on their likelihood of realization, as a thorough empirical investigation of VN has not been undertaken before.

The more general groups of 'All/Some/None Allowed' as well as 'Core/Coerced/Incompatible' follow a pattern of frequency that supports our more general hypotheses: CM frequency is highest where 'All' verbs are compatible, second-highest where 'Some' verbs are compatible, and lowest where 'None' of the verbs are compatible. Additionally, CM frequency is highest for classes where CM was thought to be 'Core,' and lower where CM was thought to be compatible only through coercion. These results are consistent with those obtained in the previous examination of CM (Bonial et al., 2011c), automatically detected in the CHILDES corpus (MacWhinney, 2000), and manually annotated in a smaller segment of the Wall Street Journal. Overall, our results from SemLink, as well as our earlier results in previous research, demonstrate the validity of the manual annotations. Table 5 gives the results of our findings, organized this time according to these more general groups.

| Grouping | CM counts | Total Frames | CM Frequency |
|---|---|---|---|
| All compatible | 3677 | 35398 | 0.1039 |
| Some compatible | 2571 | 32906 | 0.0781 |
| None compatible | 777 | 12717 | 0.0611 |
| CM is core | 4990 | 48396 | 0.1031 |
| CM coerced | 1258 | 19908 | 0.0632 |

Table 5: Frequencies of CM, regrouped according to 'all/some/none' classification and 'core/coerced' classification.

Notably, the frequency of CM instances for those classes where manual annotations found verb members to be incompatible with CM is higher than expected. This indicated that there may be error in our data. Thus, we randomly selected instances for manual inspection to ensure that our search returned appropriate instances. We found that although the overall trends seem reliable, there were errors in the form of false positives. The most predominate type of these errors are usages where the final PP is characterized by the preposition 'into', but the argument corresponds to a result or product rather than a destination. Thus, further refinement is needed as we move forward with empirical validation of VN.

### 3.1. SemLink Coverage of VerbNet

When weighing the import of these results, a key issue is the extent to which SemLink provides data for the verbs contained in VN. For us to fully understand the behavior of verbs in VN, we need to see how each verb behaves in actual usage. However, not every verb found in VN occurs in the SemLink data. Table 6 summarizes the percentage of verb types that are represented in VN across the manual annotation groups. This summary shows that, on average, only about one third of the verbs currently in VN are represented in the SemLink Data. This indicates that a full empirical validation would require additional corpora, and shows that SemLink should be expanded in order to increase its utility.

| Grouping | % Coverage |
|---|---|
| 1 - CM rep. in VN/ All/ Core | 39 |
| 2 - CM rep. in VN/ All/ Coerced | 34 |
| 3 - CM rep. in VN/ Some/ Core | 44 |
| 4 - CM not rep. in VN/ All/ Core | 32 |
| 5 - CM not rep. in VN/ All/ Coerced | 28 |
| 6 - CM not rep. in VN/ Some/ Core | 31 |
| 7 - CM not rep. in VN/ Some/ Coerced | 32 |
| 8 - CM not rep. in VN/ None | 29 |
| All Allowed | 34 |
| Some Allowed | 32 |
| Core | 35 |
| Coerced | 31 |

Table 6: Percentage of verb types in grouping that are found in SemLink data, grouped by CM manual annotation types

## 4. Comparison to FrameNet

Given that the definition and delimitation of CM is subject to interpretation, and there is a certain amount of error in our results, we compared our findings to FrameNet's representation of CM. To do this, we simply found the corresponding VN class for each of the verbs listed in FrameNet's Caused Motion frame, which contains 74 verbs. We then calculated the number of verbs from FrameNet's Caused Motion frame in the larger groupings of VN classes determined through the manual CM annotations. Table 7 gives the resulting number of FrameNet Caused Motion frame members across the same annotation groups.

| Grouping | # of FN Verbs |
|---|---|
| 1 - CM rep. in VN/ All/ Core | 54 |
| 2 - CM rep. in VN/ All/ Coerced | 0 |
| 3 - CM rep. in VN/ Some/ Core | 0 |
| 4 - CM not rep. in VN/ All/ Core | 20 |
| 5 - CM not rep. in VN/ All/ Coerced | 0 |
| 6 - CM not rep. in VN/ Some/ Core | 0 |
| 7 - CM not rep. in VN/ Some/ Coerced | 0 |
| 8 - CM not rep. in VN/ None | 0 |
| All Allowed | 74 |
| Some Allowed | 0 |
| Core | 74 |
| Coerced | 0 |

Table 7: Counts of FrameNet verbs found in Caused Motion frame, grouped by CM manual annotation types

Although this is certainly not a comprehensive examination of where motion appears in FrameNet with an outside cause, this simple comparison validates the overall trends in our findings. The majority of the verbs in FrameNet's Caused Motion frame (54 of the total 74 verbs) fall into classes where VN also recognizes CM as part of the verb's core behavior (Group 1); this evidences the quality of VN's current representations. However, approximately one-quarter of the verbs listed in FrameNet's Caused-Motion frame map to Group 4, which is comprised of classes in VN that do not include a representation of CM. It was also this group that had the highest frequency of CM in our data, and where we had hypothesized a higher frequency of CM occurrences because the manual annotations found these verbs to be compatible with CM as part of their core behavior. Again, this demonstrates that CM should be included as typical behavior for the verbs of these classes.

## 5. Future Work and Conclusions

In order to facilitate complete empirical validation of VN, we suggest that VN's frames be made hierarchical. We are beginning to work on a resource that will map the more fine-grained frames currently found in VN to successively coarser-grained frames. The superordinate frames would be frames containing only the most basic semantic information (e.g., Agent V Theme). This mapping resource will allow users of VN, who are trying to apply or match VN's syntactic frame information to real data, to back off from the syntactic specificity of the current frames to a level of specificity that is well-suited for their needs. Users could then use the more basic frames where syntactic patterns of the data do not match the patterns expected in the current VN frames. Thus, for example, any syntactic constituents acting as Agent and Theme would match the most basic pattern, where currently syntactic specificity often requires that roles such as Agent be realized in the prototypical form of a NP. The lower levels of the frame hierarchy could become increasingly specific, populating syntactic constituents and, where practical, particularly informative lexical items, such as 'that,' which VN currently includes.

Although VN is somewhat hierarchical, as mentioned in the background to VN, what we are proposing is making the frames themselves hierarchical. Currently, VN's classes are only hierarchical in the sense that lower levels of classes contain additional frames that are compatible with only a subset of verbs in the class. Therefore, although each subclass does add to the number of syntactic frames that verbs within the subclass are compatible with, the syntactic complexity of the frames themselves remains constant across the top level of each class as well as its subclasses. In some cases, as discussed previously, this level of complexity is often detailed with so much syntactic specificity that it is very difficult to find instances that match the frames perfectly in real data.

The utility of hierarchies in facilitating ease of interoperability and improving the overall user experience of VN was similarly demonstrated in the recent development of a thematic role hierarchy, soon to be released with a new version of VN (Bonial et al., 2011a; Bonial et al., 2011b). This hierarchy was the result of a systematic comparison of the VN thematic roleset to that of LIRICS.[4] LIRICS is another semantic resource created with the aim of establishing sets of annotation concepts, defined in accordance with the International Organization for Standardization (ISO) standard 12620 as data categories, for syntactic, morphosyntactic, and semantic annotation. Through this comparison,

[4] Linguistic InfRastructure for Interoperable ResourCes and Systems `http://LIRICS.loria.fr`

we found that the VN thematic roleset was comparatively fine-grained, with some thematic roles that are specific to certain types of events or verbs classes. Although the fine-grained roles may be helpful for distinguishing characteristics of verb classes, some users may find it advantageous to have a thematic roleset that is more general and applicable to all event types, more like that of LIRICS. Additionally, we found that having both fine-grained and coarse grained roles allows for the VN roleset to map more easily to more fine-grained semantic resources, such as FrameNet, as well as more coarse-grained semantic resources, such as PropBank. With the aid of a thematic role hierarchy, users can select the level of granularity that is ideally suited to their task simply by selecting the appropriate level of the the hierarchy. The intermediate level of this hierarchy largely overlaps with the LIRICS roleset, and seems to be the level that is maximally descriptive while also generalizable to all types of events. Thus, this level has served as a starting point in another facet of our ongoing research, the potential creation of an ISO standard set of roles.

As previous research has suggested (Bonial et al., 2011c), we also find that it would be beneficial for VN classes to be augmented with probabilities of verbs being realized in a particular syntactic frame. Given that verbs can constantly be used in new ways, it would be informative for frames to be empirically generated and listed with a probability of realization, drawn from a variety of corpora. An important advantage would be that coercive usages could be included in VN, where currently such novel, creative usages of language aren't accounted for in the resource.

Our research has demonstrated that there are certain gaps in how verbal behavior is represented in VN. The mapping resource we have suggested will enhance the interoperability of VN, while also allowing us to complete an empirical validation of VN and gather probabilistic information about the likelihood of each realization. We suggest a final step of incorporating such probabilities into VN and adding frames where we find evidence that a class frequently participates in that frame. Ideally, we could easily update this information by continuing to compare the class representation against additional corpora, such that VN would serve as a model of ever-changing language.

## 6. Acknowledgements

We gratefully acknowledge the support of the National Science Foundation Grant NSF-IIS-1116782, A Bayesian Approach to Dynamic Lexical Resources for Flexible Language Processing and DARPA/IPTO funding under the GALE program, DARPA/CMO Contract No. HR0011-06-C-0022, VerbNet Supplement. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## 7. References

Bonial, C., Brown, S.W., Corvey, W., Palmer, M., Petukhova, V., and Bunt, H. 2011a. An exploratory comparison of thematic roles in verbnet and lirics. In *Sixth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6)*, pages 39–44.

Bonial, C., Corvey, W., Palmer, M., Petukhova, V., and Bunt, H. 2011b. A hierarchical unification of lirics and verbnet semantic roles. In *Proceedings of the ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ICSC 2011)*.

Bonial, C., Windisch Brown, S., Hwang, J. D., Parisien, C., Palmer, M., and Stevenson, S. 2011c. Incorporating coercive constructions into a verb lexicon. In *ACL 2011 Workshop on Relational Models of Semantics held in conjunction with ACL-2011*, Portland, Oregon, June.

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. L. 2002. Background to framenet. *International Journal of Lexicography*, 16(3):235–250.

Goldberg, A. E. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Joanis, E., Stevenson, S., and James, D. 2008. A general feature space for automatic verb classification. *Natural Language Engineering*, 14(3):337–367.

Kipper, K., Korhonen, A., Ryant, N., and Palmer, M. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation Journal*, 42:21–40.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

Loper, E., Yi, S., and Palmer, M. 2007. Combining lexical resources: Mapping between propbank and verbnet. In *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS-7)*, Tilburg.

MacWhinney, B. 2000. The childes project: Tools for analyzing talk. *The Database*, 2.

Marcus, M., Santorini, B., and Marcinkiewicz, M. A. 1993. Building a large annotated corpus of english: the penn treebank. *Computational Linguistics*, 19(2):257–285.

Palmer, M., Gildea, D., and Kingsbury, P. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105.

Palmer, M. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference, GenLex-09*, Pisa, Italy.

Swier, R. and Stevenson, S. 2004. Unsupervised semantic role labeling. In *Proceedings of the 2004 Conf. on Empirical Methods in Natural Language Processing*, pages 95–102, Barcelona, Spain.

Zaenen, A., Condoravdi, C., and Bobrow, D. G. 2008. The encoding of lexical implications in verbnet. In *Proceedings of LREC 2008*, Morocco, May.

# A Universal Representation for Shallow and Deep Semantics

**Mehdi Manshadi[1], James Allen[1,2]**

[1]Department of Computer Science, University of Rochester Rochester, NY
[2]Institute for Human and Machine Cognition, Pensacola, FL
{mehdih,james}@cs.rochester.edu

**Abstract**

We define a graphical semantic representation that readily captures the partial semantic analyses produced by shallow processing techniques, yet is also as fully expressive as the representations used in deep analysis systems, including discourse processing. While in most existing natural language systems, robustness often comes at the expense of shallowness, our representation is designed to bridge this gap. The framework is not specific to a particular semantic theory, and may be translated into various target languages. In particular, the translation into first order or intentional logic is transparent. We show how the framework is able to capture more complex semantic phenomena, such as scopal adverbials and predicate modifiers. The graphical frameworks allows us to define mathematical notions to determine the well-formedness of a representation or the coherence of the corresponding sentence once we have the complete semantic representation of a sentence. A unique property of our semantic framework is to encode some syntactic properties of a sentence as well. We define an evaluation framework for this formalism that allows one to compute semantic recall and precision measures given gold standard representations.

**Keywords:** Semantic Representation, Semantic Evaluation, Underspecification, Robust Semantics

## 1. Introduction

There is a growing interest in semantic interoperability to enable research on using and combining semantic knowledge form different sources for deeper language understanding. But most previous semantic resources have focused on only one particular aspect of semantics, e.g., word senses, semantic roles, coreference, dependencies, etc. And other areas, such as scoping of quantifiers and operators, have been little studied. While deep understanding has been mostly the focus of symbolic NLP over the past two decades, in the last few years there has been some effort to semantically interpret text using statistical methods (Zettlemoyer and Collins, 2009; Clarke et al., 2010; Vogel and Jurafsky, 2010; Liang et al., 2011; Chen and Mooney, 2011; Branavan et al., 2011). Those works, however, often define their own semantic representation. This paper is an effort to suggest a unified framework that could be useful to both statistical and symbolic approaches because it creates a semantic representation from a process as shallow as POS tagging and goes all the way to scope disambiguation.

Thus the goal here is to present a formalism that would allow all such information to be combined into a single representation to enable sharing resources more easily, and to make a significant step towards building semantic resources that provide deeper semantic information than previously possible.

Key requirements on such a representation are

1) **Incrementality** - the representation should be able to store partial representations (e.g., just word senses, or word senses and semantic roles), and facilitate adding other aspects of semantics at a later date

2) **Interoperability** - the representation should be able to extract partial semantic representations from existing formats and regenerate such formats from full semantic representations

3) **Expressibility** - the representation should be able to express to the best of our abilities, the subtleties and phenomena captured in current state of the art models of natural language semantics (e.g., Copestake et al., 2005).

These points reflect the fact that our goal is a pragmatic one. The first two requirements above specify that the formalism should let us bootstrap a rich semantic resource by combining analyses by existing techniques that each address part of the problem. The third requirement guarantees that the representation serves the need of future work as research moves to deeper semantic processing. It requires "head room" in the representation, making sure the formalisms is as expressive as the best state-of-the-art constraint-based semantic representations.

In addition, to be a useful formalism for semantic interoperability, we add two additional requirements:

4) **Readability** - the representation should be relatively easy to browse and understand by humans

5) **Evaluation** - the representation should support precise evaluation metrics that can be applied to work at different

levels, whether it just addresses a single aspect such as word sense disambiguation, or is used for evaluating systems that attempt a full deep semantic analysis.

Its important to distinguish two separable problems in semantic evaluation, namely the ontology used and the structure of the logical form language (LFL). The ontology determines the set of word senses and semantic relations that can be used. The LFL determines how these elements can be structured to capture the subtle meanings of natural language sentences. We are focusing on the latter in the paper, the logical form language, and our framework would work with any ontology.

The formalism is graphical in nature. Besides provide a framework for capturing the output of a wide range of systems, the graphical structure provides an intuitively readable semantic formalism, one that we find much easier to read than, say, an equivalent expression of the same content in a logical language. Gaps in analyses and connections between terms are readily visible in graph form. In addition, the graphical framework facilitates the construction, semiautomatically with human editing, of gold-standard representations.

The evaluation framework is couched in graph matching and we present an algorithm for rapidly computing node alignments that maximize the score (heuristically, as the problem in general is NP-hard). Because of the nature of the representation, a single gold-standard for a deep logical form can be used to evaluate the outputs from a wide range of different systems, ranging from word sense disambiguation and semantic role labeling, to complex semantic phenomena such as adverbials, predicate modifiers and quantifier scoping, to some discourse phenomena (e.g., coreference resolution).

## 2. The LF Graph Framework

LF graphs are designed to be an expressive, yet intuitive, formalism for expressing sentential logical form. In designing the LF, we wanted a formalism that first, could integrate information from different processing steps such as part-of-speech tagging, word sense disambiguation, NP chunking, etc., to form a deep semantic representation, and second, could readily express partial analyses of utterances from a deep representation in order to support evaluation of a wide range of techniques from statistical methods (semantic role labeling, dialog act tagging) to deep parsing producing detailed logical forms. But to provide headroom for future work, it also needs to be expressive, providing good coverage of the complex semantic phenomena in language, including modal operators, generalized quantifiers, and underspecified scoping constraints (cf Minimal Recursion Semantics (MRS); Copestake et al., 2001 & 2005).

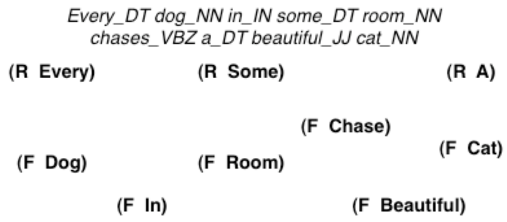


*Figure 1. Semantic representation using POS tags*

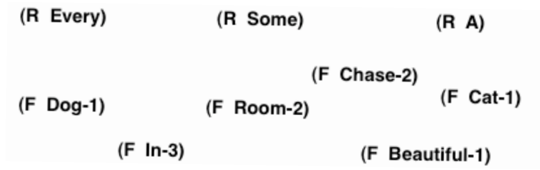


*Figure 2. Adding word senses*

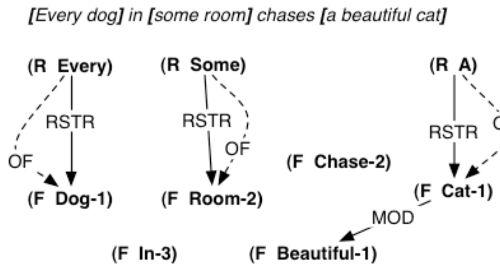


*Figure 3. Adding information from NP chunking*

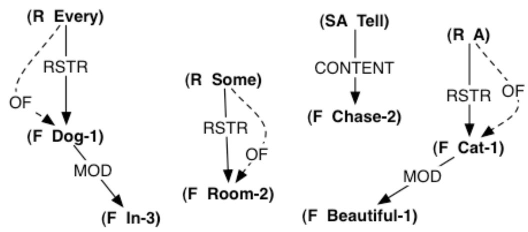


*Figure 4. Adding information from parse tree*

We introduce the LF graph by demonstrating the process of building this deep representation from the output of the shallow processing techniques. Consider the sentence *Every dog in some room chases a beautiful cat*, and let's run a part-of-speech (POS) tagger on it. We create a semantic representation only by using the POS tags as shown in Figure 1. All the DT tags create **R-nodes** in the LF, where R is for **referent** and shows that an entity has been introduced in the domain of discourse. The NN, VBZ, and JJ nodes create a second type of node, the **F-nodes**. F-nodes are **formula** nodes representing first order predicates. Now assume that we run a WSD module on the sentence to obtain the sense of each word. This helps to further specify the F-nodes by assigning the actual predicate (concept) from an ontology instead of a prototype predicate represented by the word itself as shown in Figure 2. Let's continue the process of applying shallow processing techniques to the sentence. This time we run an NP chunker, as shown in Figure 3. The NP chunker helps us to deepen the representation by introducing two kinds of edge. The **solid** edges are the **structural** edges. They show how the nodes are combined to form logical formulas. For example, the edge labeled *RSTR* shows that the restriction of the quantifier *Every* is the predicate *Dog*. The **dotted** edges are **argument** edges, representing the arguments of the predicates. For example the dotted edge labeled *Of,* from *Every* to *Dog,* shows that the entity introduced by the quantifier *Every* is the argument of the predicate *Dog*. The direction of the argument edges may seem counter -intuitive. The direction is chosen to represent outscoping relations, that is the dotted arrow from *(R Every)* to *(F Dog)* shows that this predicate must be bound within this quantifier. This direction will be consistent with the direction of dominance (i.e. outscoping) constraints we define on the scope restrictions later in this section. Let *x* be the entity introduced by the node *(R Every),* then the two nodes with the edges between them (Figure 3) could be represented as the following logical formula:

1. *Every(x, Dog(x), ...)*

where the body of the quantifier *Every* is not specified. If we take a neo-Davidsonian approach, and let *f* be the variable introduced by the node *(F Dog)*, we can obtain the following logical formula:

2. *Exists(f, Dog(f), Every(x, Of(f,x), ...))*

In general the transformation of the graph into a logical formula is transparent. Every R-node introduces a first order variable, and every F-node introduces a variable representing the reification of a predicate. Every F-node *(F Pred)* with incoming argument edges $Arg_1$, $Arg_2$, etc., is transformed into a conjunction of predicates:

3. $Pred(f) \wedge Arg_1(f, x_1) \wedge Arg_2(f, x_2) \wedge ...$

where $f$ is the variable introduced by the node *(F Pred)* and $x_1$ .$x_2$, etc. are the variables introduced by the head node of $Arg_1$, $Arg_2$, etc.

Going one step deeper, let's assume that we have the syntactic tree of the sentence, where the preposition phrase is attached to the NP *Every dog* and *Chases* is the head of the sentence. By incorporating the syntactic tree, we can form the semantic representation in Figure 4. This figure shows a third type of node the **speech act** or **SA-node.** The SA-node specifies the speech act of the utterance, and is connected to the main predicate of the sentence by an edge labeled *CONTENT.* If we ignore the dotted edges in this figure, it is formed of 4 trees, 3 of them corresponding to the three shallow NPs and one corresponding to the main predicate of the sentence, which we call the **heart** formula. This holds in general. That is every parse tree can be transformed into a forest of exactly *n*+1 (*n*: the number of NPs) trees rooted at the *n* R-

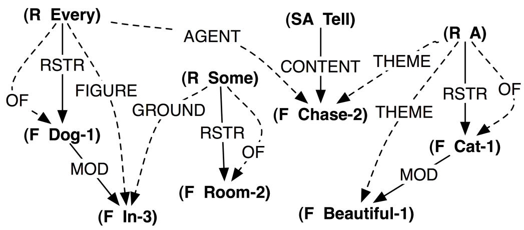


*Figure 5. Adding semantic roles*

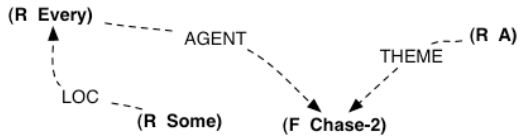


*Figure 6. Heart-connectedness*

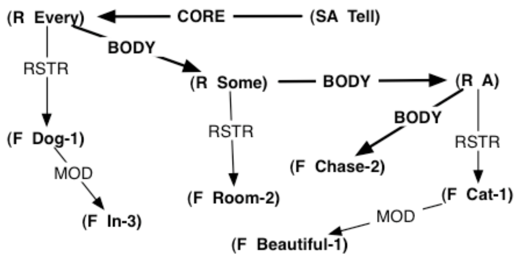


*Figure 7. Scope disambiguation (dotted edges were removed for clarity)*

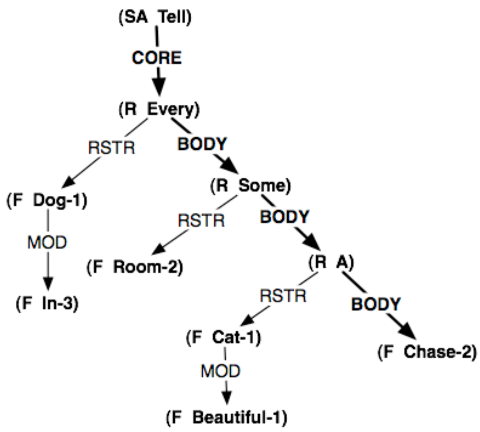


*Figure 8. A fully scope disambiguated tree structure*

nodes and a single SA-node (Manshadi et al., 2008). This is called **canonical form** (CF). Note that CF is scope underspecified that is the body of the quantifiers are left **underspecified**.

Going deeper in the analysis of the sentence, a comprehensive semantic role labeling (SRL) module can specify the complete argument structure of the predicates as shown in Figure 5. By comprehensive SRL, we mean a module that not only tags the role of the verb predicates, as in PropBank (Palmer et al., 2005), but also determines the semantic relations for nouns, as in NomBank (Meyers et al., 2004).

If we go further in the level of underspecification allowed and for example allow the label of the argument edges to be left underspecified, more information could be represented by shallower processing steps. For example, only looking at the parse tree, we can infer that there are argument edges between the nodes *(R Every) /(R A)* and the node *(F Chase)*, although we leave it to the SRL to specify the label of these edges.

An interesting property of a canonical form with complete argument edges is that given a **coherent** sentence, if we collapse the trees formed by solid edges into a single node, as shown in Figure 6, the resulting graph is **heart-connected**, that is every node reaches the heart by a directed path. This gives a mathematical characterization of **coherence**. Intuitively, this holds because in a coherent sentence every shallow NP must contribute to the overall meaning of the sentence either by directly being an argument of the heart formula (i.e. reaching heart by a directed path of length 1), or by modifying an argument of the heart formula (i.e. reaching heart by a directed path of length 2), and so on. The detailed justification of this definition is beyond the scope of this paper.

The representation can still go deeper. The semantic representation in Figure 5 is scope underspecified, that is the quantifier scope preference is not yet determined. The quantifier scoping can be specified simply by adding body edges, another type of structure edge. This has been shown in Figure 7. Note that every fully scope-disambiguated representation forms a tree (ignoring the doted edges). If we rearrange the nodes in Figure 7 we will have the tree in Figure 8.

## 3. Scope Underspecification

Another feature of our representation is that scope restrictions can be readily added to the representation with dotted edges without inventing a new mechanism. For example a dotted edge from *(R Every)* to *(R Some)* (Figure 9) means that in every scope-resolved version of this representation the node *(R Every)* must outscope the node *(R Some)*. Interestingly this is a property that all other dotted edges implicitly carry. This is because dotted edges represent arguments of the predicates hence express the binding constraints, and binding constraints are nothing but outscoping (or **dominance**) relations which force the head of a dotted edge to outscope (dominate) its tail in every scope-resolved structure. In fact, if we ignore the label of the edges (Figure 9), our representation is an underspecified structure in the framework of Dominance Constraints (Egg et al., 2001). Not all the constraint-based formalisms however use dominance relations as constraints. For example, MRS uses a restricted version of dominance constraint, called **qeq** (equality modulo quantifier) relation. This highlights another property of our representation. If a quantifier *Q2* occurs in the restriction of a quantifier *Q1*, a

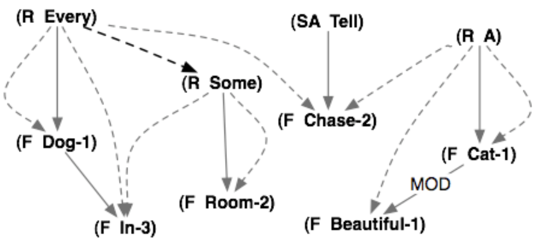


*Figure 9. Scope restriction as dominance constraints*

qeq relation forces the restriction predicate of *Q1* to be in the body of *Q2* as for the predicate *Dog(x)* in the following logical formula for the sentence *Every dog in a room barks*.

4. *Every(x, A(y, room(y), Dog(x) ∧ In(x,y)), Bark(x))*

As mentioned above, our framework converts the syntactic tree of a complete sentence to a Canonical Form, Manshadi et al. (2008) prove that for structures in canonical form, the dominance and qeq relationships become equivalent. Therefore, we can simply use dominance constraints and still remain consistent with MRS or Robust MRS (Copestake 2007).

Finally, there is an issue of intractability with most constraint-based frameworks. Note that to build the scope-resolved structures, we have to solve the constraint-based representation; that is to find a solution that satisfies all the dominance constraints. This has been shown to be intractable for Dominance Constraints and MRS in general (Althaus et al., 2003). Here the notion of heart-connectedness will save us. Remember that every coherent natural language sentence has a heart-connected representation in our framework. Manshadi et al. (2009) prove that for every heart-connected graph, the dominance constraints can be solved in polynomial-time.

## 4. Evaluation Framework

This section defines precision and recall measures for LF graphs. Given a gold-standard LF-graph, we can evaluate the LF graph produced by a system by defining node and edge scoring criteria and then computing the node alignment that maximizes the overall score.

The evaluation metric between a gold LF graph G and a test LF graph T is defined as the maximum score produced by any node alignment from the gold to the test LF (see Figure 10). More formally, an alignment A is a one-to-one mapping from the nodes of the gold graph to nodes of the test graph (or to a pseudo empty node if there is no corresponding node in the test graph). Once we have defined a scoring metric between aligned nodes and edges, we define the match between a gold and test graph as the maximum score produced by an alignment. While more complex scoring functions can be used (e.g.. Resnik and

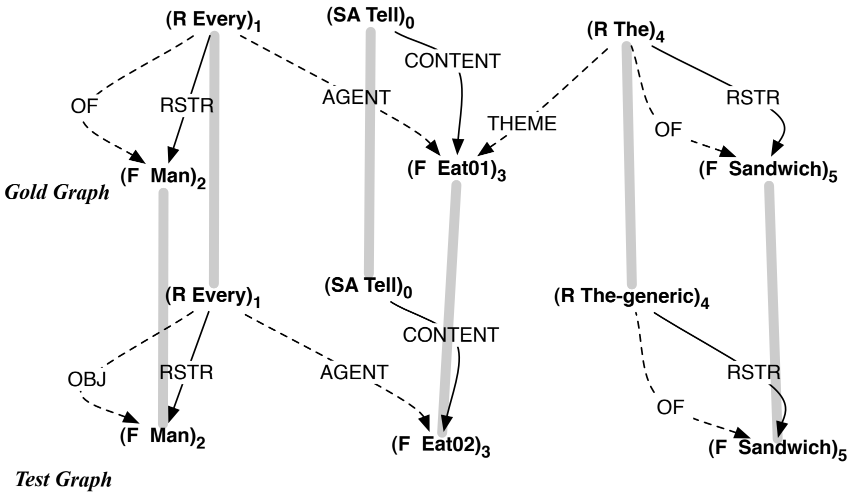


*Figure 10: A node alignment for computing precision and recall*

Yarowsky, 1997; Melamed and Resnik, 2000), for illustration purpose here we will use a very simple measure:

- *Nscore$_A$(n)* = 2 if the indices match, and both the node type and sense in the label of node *n* matches the label of node *A(n)*, 1 if one of them matches, and 0 otherwise.
- *Escore$_A$(e)* = 1 if e connect nodes *n1* and *n2*, and there is an edge between *A(n1)* and *A(n2)* with same label, 0 otherwise.

While we use a simple all-or-nothing metric on the labelled edges, more complex schemes are easily developed. We could, for instance, have a hierarchy of semantic roles, allowing very abstract roles (e.g., *A0, A1, ...)* as generalizations of the semantic roles we have used here, and give partial scores for abstract labels. And it may be that we might use different evaluation metrics based on the goals of the research.

Once we have a node and edge scoring scheme defined, we can define the overall graph match score as the score of the alignment that maximizes the sum of the node and edge scores.

5. $Gscore(G,T)=max_A(Sum_{n,e}(Nscore_A(n)+Escore_A(e))$

Once we know Gscore(G,T), we can compute semantic precision and recall measures by comparing this to the G and T graphs aligned with themselves, which gives us the maximum possible gold and test scores.

6. $Precision(G,T) = Gscore(G,T)/Gscore(T,T)$

7. $Recall(G,T) = Gscore(G,T)/Gscore(G,G)$

As an example, Figure 10 shows an alignment with a gold and a test LF graph, where the system makes several mistakes. It gets the wrong sense of the verb *eat,* fails to identify *the sandwich* as the THEME role of the verb *eat,* and it interprets *the sandwich* as a generic rather than a specific reference (as in *the sandwich was invented in 1789*). Of the six nodes, 4 match perfectly, yielding 8 points, and the other two identify the correct semantic role in the sentence (i.e., a predicate and a quantifier), but get the wrong sense. Giving 1 point each, we have a node score of 10 (out of a maximum 12). With these nodes aligned, we can

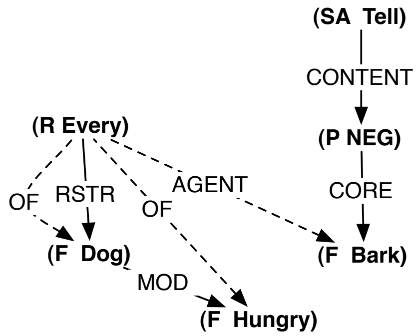


*Figure 11: The LF graph for* Every hungry dog did not bark

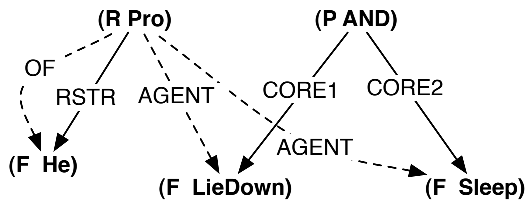


*Figure 12: Pronouns and Conjunction*

then compute the edge score. Regarding the edges, there are five exact matches, one edge mislabeled and one missing, and two spurious extras. This gives an edge score of 5, and thus *Gscore(G,T)* is 15. Computing *Gscore(G,G)*, the goal matched against itself, yields 19, and *Gscore(T,T)* yields 20. Thus we have a precision of 75% (15/20) and a recall of 79% (15/19).

Computing the best alignment between two graphs is an NP-hard problem in general, but because of the fact that a large number of node labels are unique, we have found that heuristic methods work well. Note that while one might think the node index (i.e., word position) would uniquely align nodes, this is only for the simplest cases. Because semantic structures arise from phrasal structure, not isolated words, there are cases where nodes don't correspond to a specific position. We first use the technique of similarity flooding (Melnik et al, 2002) to compute an estimated similarity score between nodes, and then a best-first heuristic search over the results to find a high-scoring alignment that is consistent overall.

## 5. Handling Other Phenomena

This section briefly considers a range of more complex semantic phenomena and shows how LF graphs can accommodate them.

### 5.1 Scopal Operators

One important phenomena that we haven't discussed yet are the scopal operators, including the logical operators such as negation and conjunction, as well as some adverbial modifiers. Similar to quantifiers, scopal operators create scopal ambiguity (cf. *fixed scopals* in Copestake et. al, 2005). Hence we define our last type of node, the **P-nodes**, to deal with them. As a simple example, Figure 11 shows the LF graph for the sentence *Every hungry dog did not bark*, which is of course ambiguous between an interpretation where *no dogs bark* (negation inside the scope of *Every*) and there being *some dogs that do not bark* (*Every* within the scope of *not*).

Conjunctive operators are treated in a similar way, but may take multiple arguments, which by convention we name CORE1, CORE2, and so on. Such constructions also introduce ellipsis, as in *He lay down and slept*, which describes a sequence of events, both by the agent *He*, shown in Figure 12. Note in this example that pronouns are treated as quantifiers, with the semantic content encoded in the pronoun as a predicate (in this case *He)*. Note also that this analysis has inserted the AGENT link from the sleep event to the pronoun, the result of basic ellipsis resolution.

### 5.2 Predicate Modifiers

Many modifiers are best treated as modifying the predicate term itself, rather than the objects being referred to. For instance, the object referred to as *A very beautiful cat*, is a *cat*, and is *beautiful*, but the *cat* itself is not *Very*! Rather, the *cat* has the property of being *very beautiful*. With properties being reified as nodes in LF graphs, predicate modifiers can be simply captured by indicating this property as the argument. Figure 13 shows the LF graph for this. Note that the OF argument of the predicate *Very* is the predicate *Beautiful*.

Note while the LF graph can capture these distinctions, it does not commit to a specific semantic interpretation for constructions. For instance, while the predicate *small* in *A small cat* is typically treated as an intersective adjective, *i.e.*, an object *x* such that *Cat(x)* ∧ *Small(x)*), many would argue it is a predicate modifier, noting that a small cat is a very different size than a small elephant. Once we start developing computational approaches to distinguish between intersective and predicate adjectives, the LF graph formalism can make this distinction.

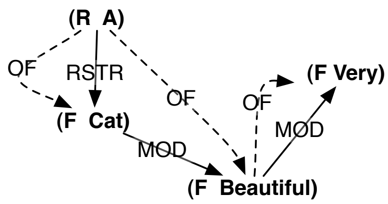


*Figure 13: LF graph showing a predicate modifier in* A very beautiful cat

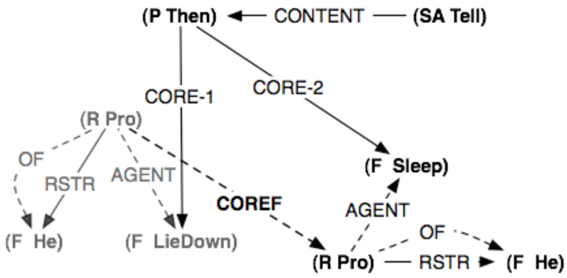


*Figure 14: Discourse connectives and coreference:* He lay down. Then he slept.

### 5.3 Discourse Connectives and Coreference

As a final example, consider discourse markers, as in the second sentence in the text fragment *He lay down. Then he slept*. The word *then* is a discourse connective that relates the events in the two sentences, not the adverbial reading that modifies the time of the sleeping (as in *He slept then*). Discourse adverbials may be treated as scopal operators as shown in Figure 14.

Finally, note that coreference can be easily captured by a new COREF link between the nodes representing the discourse entities, as shown in Figure 14.

## 6. Related Work

In the past two decades, there have been many efforts to annotate natural language sentences with semantic information. However, most of these efforts have focused on a piece of semantics such as word senses (SEMCOR; Fellbaum, 1997), semantic roles for verb predicates (PropBank; Palmer et al., 2005), semantic relations for noun predicates (NomBank; Meyers et al., 2004), discourse markers (Penn Discourse Bank; Prasad, 2008), etc. This is because focusing on a specific semantic piece, it is easier for the annotators to capture all the phenomena in that narrow area. In the past several years, there have been some efforts to put these pieces together to build a detailed semantic representations (Pustejovsky et al., 2005). While these efforts have focused on how to integrate the existing resources (by resolving the potential conflicts) in order to build a sound and detailed semantics, our work is about the representation of such detailed semantics. The raw detailed semantics obtained by integrating different resources is often a complex network of nodes and relations. We build our framework on their core idea of forming a detailed semantic representation by integrating pieces of information which are either manually or automatically created. However,, the graphical formalism with a predefined set of node types and its separation of the structural and argument edges, as well as definition of well-formedness and coherence, leads to a mathematically well-defined, fairly readable framework that can transparently be transformed into a target semantic language.

There exist some other frameworks that use graphs to represent semantic information e.g. ConceptNet (Liu and Singh, 2004). Those frameworks however are often not about detailed semantic representation of natural language sentences. ConceptNet, for example, is a graphical framework for knowledge representation. Other graphical formalisms such as the Constraint Language for Lambda Structures (Egg at al., 2001) represent fairly detailed semantic representation of natural language, but do not address robustness, incrementality, and granularity, the main properties of our framework.

Robust Minimal Recursion Semantics or RMRS (Copestake, 2007) is probably the closest work to ours. While some of the ideas of our framework have been taken from this formalism, our framework has major advantages over RMRS:

- While RMRS is predicate logic, ours offers a graphical form, an intermediate representation that can be translated into various target semantic languages.
- The canonical form, which we adopted, to some extent captures the syntax of a sentence. For example, corresponding to each shallow NP there is one tree (composed of structural links), rooted at the quantifier/determiner of that NP (if it has any, otherwise a bare R-node). On the other hand, the graph can transparently be converted to a target semantic language. Therefore a single representation captures both syntactic and semantic information of a sentence. That is a significant advantage when it comes to the syntactic ambiguity, because instead of having multiple syntactic trees, the ambiguity can be left underspecified in a single LF graph.
- Our framework supports different levels of granularity, therefore for applications where a coarse-grained semantic representation is adequate, we don't have to deal with complex semantic analysis, or the ambiguities that may result from such fine-grained analysis.
- The heart-connectedness property gives a clean mathematical criteria for the coherence of a semantic representation.
- As a scope underspecification formalism, our framework is genuinely constraint-based, therefore it remains true to the incrementality principle even when it comes to the quantifier scoping. To further explain this, remember that MRS uses qeq relations which are a restricted version of outscoping constraints. Those constraints can only model the relation between a quantifier and its restriction predicate (or conjunction of predicates), therefore in practice qeq relations cannot be used to put further constraints on the permissible readings (e.g., to force island constraints). Our framework, on the other hand, uses outscoping constraints. Although equivalent to qeq relations at the syntax/semantic level, outscoping constraints can be further (that is after syntax/semantic interface) added to the semantic representation to filter out unwanted readings (possibly by deeper processing levels such as discourse and/or pragmatic knowledge). In addition, the scope constraints naturally fit into our framework, and no new mechanism is invented to handle those constraints.

RMRS, on the other hand, has mechanisms for underspecification of word senses and argument structure that we haven't explored yet.

## 7. Summary

We have described a semantic formalism that is suited for capturing the output of a wide range of semantic formalisms, from word sense disambiguation tasks and semantic role labeling, to some aspects of discourse processing. The formalism is notable in that each level of complexity can be added incrementally. As a result, the representation could be promising as a framework in which we can compare the results from different frameworks. We have presented an evaluation metric to measure the precision/recall of detailed or partial semantics of natural language sentences represented in our framework, in order to compare two given NLP systems, or an NLP system vs gold standard annotation.

Because of its incremental nature, our formalism is useful as a formalism for studying direct incremental construction of semantic representations from text. In the past several years, there has been a great interest in adopting such approaches for developing end to end systems that require natural language understanding, such as answering natural language queries to a geography database, finding directions following natural language instructions, etc. We hope that the robust incremental framework presented here encourages those efforts to adopt such a uniform representation in order to be able to share resources, and more importantly, to easily adapt such systems to new domains.

## 1 Acknowledgements

This work was support in part by NSF grants 0958193 and 1012205, and ONT grant N000141110417.

## References


Althaus, E., Duchier, D., Koller, A., Mehlhorn, K., Niehren, J., and Thiel S. 2003. An efficient graph algorithm for dominance constraints. Journal of Algorithms, 48:194–219.

Branavan, S. R. K., Silver, D., Barzilay, R., 2011. Learning to win by reading manuals in a monte-carlo framework. In ACL, pages 268–277.

Chen, D. L. and Mooney, R. J. 2011. Learning to interpret natural language navigation instructions from observations. In AAAI.

Clarke, J., Goldwasser, D., Chang, M., and Roth, D., 2010. Driving semantic parsing from the world's response. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010), pages 18–27, Uppsala, Sweden.

Copestake, A., Flickinger, D., Pollard, C., and Sag, I. 2005. Minimal Recursion Semantics: An Introduction. Research on Language and Computation, 3(4):281-332.

Copestake, A. 2007. Semantic composition with (robust) minimal recursion semantics. In ACL-07 work- shop on Deep Linguistic Processing, Prague.

Copestake, A., Lascarides, A. and Flickinger, D. 2001. An Algebra for Semantic Construction in Constraint-Based Grammars. ACL-01. Toulouse, France.

Davidson, D., 1967. The logical form of action sentences. In Nicholas Rescher, (Ed.), The Logic of Decision and Action, pp. 81–95. U. of Pittsburgh Press.

Egg M., Koller A., and Niehren J. 2001. The constraint language for lambda structures. Journal of Logic, Language, and Information, 10:457–485.

Fellbaum, C., (1997), WordNet: An Electronic Lexical Database, MIT Press..

Koller, A. and A. Lascarides, 2009. A Logic of Semantic Representations for Shallow Parsing. Proc. 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp 45-459, Athens.

Liang, P., Jordan, M. I. and Klein, D., 2011. Learning dependency-based compositional semantics. In ACL, pages 590–599.

Liu, H. and Singh, P., 2004. ConceptNet: A Practical Commonsense Reasoning Tool-Kit. BT Technology Journal, Volume 22, Issue 4, October 2004.

Manshadi, M., Allen J., and Swift, M. 2009. An Efficient Enumeration Algorithm for Canonical Form Underspecified Semantic Representations. Proceedings of the 14th Conference on Formal Grammar (FG 2009), Bordeaux, France July 25-26, 2009.

Manshadi, M., Allen J., and Swift, M. 2008. Toward a Universal Underspecified Semantic Representation. Proceedings of the 13th Conference on Formal Grammar (FG 2008), Hamburg, Germany.

Melamed, D., and P. Resnik 2000. Tagger Evaluation Given Hierarchical Tag Sets, Computers and the Humanities 34(1-2).

Melnik, S., H. Garcia-Molina and E. Rahm 2002. Similarity flooding: a versatile graph matching algorithm and its application to schema matching, 18th Interna- tional Conference on Data Engineering,. 117-128.

Meyers, A., R. Reeves, C. Macleod, R. Szekely, et al., (2004) The NomBank project: An interim report, HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation, 2004

Palmer, M., Gildea, D., Kingsbury, P., 2005. The Proposition Bank: A Corpus Annotated with Semantic Roles, Computational Linguistics Journal, 31:1, 2005.

Pradhan, S., E. Loper, D. Dligach and M. Palmer, SemEval-2007 Task-17: English Lexical Sample, SRL and All Words 2007. in Proc Fourth International Workshop on Semantic Evaluations (SemEval-2007).

Prasad, R., Dinesh, N., Lee, A., Milt-Sakaki, E., Robaldo, L., Joshi, A., and Webber, B.. 2008. The Penn Discourse TreeBank 2.0. Proc. 6th International Conference on Language Resources and Evaluation (LREC).

Pustejovsky, J., Meyers, A., Palmer, M., Poesio, M., 2005. Merging PropBank, NomBank, TimeBank, Penn Discourse Treebank and Coreference. Proceedings of the Workshop on Frontiers in Corpus Annotations II.

Resnik, P. and D. Yarowsky. 1997. A Perspective on Word Sense Disambiguation Methods and their Evaluation. SIGLEX Workshop on Tagging Text, ANLP-97.

Stolcke, A., K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. In Computational Linguistics 26(3), 339–373.

Vogel, A., and D. Jurafsky. 2010. Learning to follow navigational directions. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 806–814.

Zettlemoyer, L. S. and M. Collins. 2009. Learning context-dependent mappings from sentences to logical form. In ACL/AFNLP, pages 976–984.

# First steps towards an ISO standard for annotating discourse relations

**Harry Bunt*, Rashmi Prasad** and Aravind Joshi*****

*Tilburg Center for Cognition and Communication, Tilburg University, The Netherlands
**Department of Health Informatics and Administration, University of Wisconsin-Milwaukee, Milwaukee, USA
***Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA
harry.bunt@uvt.nl, prasadr@uwm.edu, joshi@seas.upenn.edu

**Abstract**

This paper describes initial studies in the context of a new effort within ISO to design an international standard for the annotation of discourse with semantic relations that are important for its coherence, "discourse relations". This effort takes the Penn Discourse Treebank (PDTB) as its starting point, and applies a methodology for defining semantic annotation languages which distinguishes an abstract syntax, defining annotation structures as set-theoretical constructs, a concrete syntax, that defines a reference XML-based format for representing annotation structures, and a formal semantics. A first attempt is described to formulate an abstract syntax and a concrete syntax for the annotation scheme underlying the PDTB. The abstract syntax clearly shows an overall structure for a general-purpose standard for annotating discourse relations, while the resulting concrete syntax is much more readable and semantically transparent than the original format. Moreover, some additional elements are introduced which have an optional status, making the proposed representation format compatible not only with the PDTB but also with other approaches.

## 1. Introduction

With the recent availability of various types of linguistically annotated corpora developed for natural language processing (NLP), there is now an urgent need for addressing the demands for their *representational compatibility*, in order to ensure that each of these resources can be effectively merged, compared and manipulated with common software. An excellent example of the need for compatibility can be seen in the several different layers of annotations done on the Wall Street Journal (WSJ) corpus, such as POS tagging, syntactic constituency, coreference, semantic role labeling, events, and discourse relations. Although these annotations at different layers have resulted in a highly linguistically enriched corpus, efficient use of the resource for empirical NLP has been hindered by challenges in merging the linguistic data from the different levels because of their incompatible representations.

In addition to annotation representation, it is also necessary to ensure that when the same linguistic phenomenon is being annotated across different projects, each targeting a different language, domain, genre, or source text within the same genre, that this collective subcommunity agree on an *annotation schema standard* for the phenomenon. While agreement on schema standards is highly challenging to achieve, since it must be general enough to account for the full breadth of variation found across languages, domains, and genres, it is nevertheless necessary if we want to effectively utilize the collective resources for each phenomenon and move the state-of-the-art forward with big strides.

This work forms part of ISO efforts to establish international standards for semantic annotation. Two parts of the standard have so far been completed: ISO 24617-1 (Semantic annotation framework, Part 1: Time and events) and ISO 24617-2 (Semantic annotation framework, Part 2: Dialogue acts). Part 8, concerned with *relations in discourse*, was launched in 2011 and results from an agreement between the PDTB Research Group (http://www.seas.upenn.edu/~pdtb) and the ISO Working Group, ISO/TC 37/SC 4/WG 2 "Language resource management, Annotation and representation schemes", that a joint activity should take place to design an international standard for the annotation of discourse relations, taking the PDTB annotation scheme and guidelines (PDTB Group, 2008; Prasad et al, 2008) as the starting point. This work should include:

1. Adaptation of the PDTB annotation scheme as needed to conform to the requirements of ISO international standards;
2. Verification of the annotation scheme across a wide variety of languages, domains, and genres.

This paper describes preliminary studies for the first of these steps, in continuation of the work in Ide et al (2011). This part of ISO 24617 will provide definitions and representations of concepts for annotating explicit and implicit discourse relations. A notable feature of the abstract representation for the scheme is that it is designed to be flexible, to accommodate a certain degree of variation between schemes. This is implemented by means of optionality in the representation. Some novel concepts and structures are also introduced that are not represented in the current version of the PDTB.

## 2. The PDTB: A theory-neutral and lexically-grounded approach

The primary reason for adopting the PDTB as the basis for a discourse relation standard is that the framework avoids biasing the annotation towards any particular theory, and instead specifies discourse relations at a "low level" that is clearly defined and well understood. In particular, each relation, along with its two arguments, is annotated independently of other relations, and no further dependencies are shown among the relations. Thus, the argument structures annotated are strictly local. Since there is currently little agreement on a general theory of high-level discourse structure representation, with the proposed structures being variously trees, graphs, or DAGs (e.g., Hobbs, 1985;

Polanyi, 1987; Mann and Thompson, 1988; Webber et al., 2003; Asher and Lascarides, 2003; Wolf and Gibson, 2005; Lee et al., 2008) the theory-neutral approach of the PDTB should hold appeal for researchers across these theories, allowing for validation studies of the theories. In this sense, the PDTB framework provides a basis for an emergent and data-driven theory of discourse structure.

Another major appeal of the PDTB is its lexically-grounded approach to the annotation, lending to greater reliability of annotation, especially since it inferences at the level of discourse are much harder than at the sentence level.

The second (current) version of the PDTB, PDTB-2.0, is distributed through the Linguistic Data Consortium (LDC).[1]

## 3. Scope and Basic Concepts of PDTB

Discourse relations, such as causal, contrastive, and temporal relations, are considered to be semantic relations between abstract objects (eventualities and propositions), which are the arguments of the relation. The PDTB provides annotations of discourse relations, along with their arguments, senses and attributions, on the entire PTB-II portion of the WSJ corpus (Marcus, 1993), consisting of approximately 1 million words. In the rest of this section, we detail the basic concepts and elements of the PDTB annotation framework that underlie the proposed standard in this paper. It should be noted that the standards proposed here do not say anything about the overall annotation task design, workflows, and evaluation methods, for which the reader is referred to the PDTB reports and publications related to the annotation (Miltsakaki et al, 2004; Prasad et al., 2007; Miltsakaki et al., 2008; Prasad et al, 2008; PDTB-Group, 2008).

### 3.1. Discourse relations and their arguments

Discourse relations are often triggered by explicit words or phrases, such as the underlined expressions in (1a) and (1c), but they can also be implicit, as in (1b). Explicit realizations can occur via grammatically defined *(explicit) connectives* (1a), such as (subordinating and coordinating) conjunctions, adverbs and prepositional phrases, or with other expressions not so grammatically well-defined, called *Alternative lexicalizations* (*AltLex*) (1c). Each discourse relation is assumed to hold between two and only two abstract object (AO) arguments. Since there are no generally accepted abstract semantic categories for characterizing the arguments of discourse relations, they are simply labeled Arg1 (shown in italics) and Arg2 (shown in bold). For explicit connectives, Arg2 is the argument to which the connective is syntactically bound; Arg1 is the other argument.

(1) a. *Big buyers like P&G say there are other spots on the globe, and in India, where the seed could be grown* (...) <u>But</u> **no one as made a serious effort to transplant the crop.**

b. *Some have raised their cash positions to record levels.* <u>Implicit=because</u> **High cash positions help buffer a fund when the market falls.**

c. *But a strong level of investor withdrawal is much more unlikely this time around,* `fund managers said`. **<u>A major reason</u> is that investors already have sharply scaled back their purchases of stock funds since Black Monday.**

d. *Pierre Vinken, (...) will join the board as a nonexecutive director Nov. 29.* <u>EntRel</u> **Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group.**

e. *Jacobs is an international engineering and construction concern.* <u>NoRel</u> **Total capital investment at the site could be as much as $400 million**

Between two adjacent sentences not related by an explicit connective or AltLex, an implicit discourse relation can be inferred, in which case the annotator has to *insert* a connective to express the inferred relation, such as the implicit connective *because* inserted in (1b). For such *(implicit) connectives*, the labels Arg1 and Arg2 reflect the linear order of the arguments (Arg1 occurs before Arg2).

Arguments of explicit connectives can be located anywhere in the text, whereas arguments of implicit connectives and AltLex must be adjacent. For either of these, there are no syntactic constraints to how far an argument can extend. Thus, arguments can be single clauses, sentences, or multiple clauses or sentences. From a semantic point of view, however, an argument must contain the *minimal* amount of text that is required for interpreting the relation. To facilitate the minimality-driven argument annotation, arguments are allowed to have *supplementary* text associated with them. A supplementary text annotated for an argument — Sup1 for Arg1 and Sup2 for Arg2 — indicates that this text was perceived as relevant (but not necessary) to the interpretation of the argument. Example 2(a) shows a Sup2 annotation (enclosed in square brackets) from the PDTB, where the explanation provided for the "suing" is considered to be relevant to Arg2 but not *necessary* to interpret the temporal relation expressed with "then".

(2) a. *It acquired Thomas Edison's microphone patent* **and <u>then</u> immediately sued the Bell Co.** [claiming that the microphone invented by my grandfather, Emile Berliner, which had been sold to Bell for a princely $50,000, infringed upon Western Union's Edison patent.]

It is also possible for adjacent sentences in a coherent discourse to not be related by any discourse relation, in particular when the sentences are linked by an entity-based coherence relation (*EntRel*, as in (1d)), or are not related at all via adjacency (annotated as *NoRel*, shown in (1e)). Arguments of EntRel relations must be adjacent to each other and cannot contain sub-sentential spans, although they can be extended to include multiple sentences. Arguments of NoRel are like EntRel except that the adjacent sentences cannot be extended to include additional sentences.

[1] `http://www.ldc.upenn.edu`, Entry LDC2008T05.

### 3.2. Senses of discourse relations

In the PDTB, senses of discourse connectives are represented in a flexible manner, via a three-tiered hierarchical classification going from four coarse-grained senses at the top *class* level to more refined meanings at the second *type* and third *subtype* levels. The full PDTB sense hierarchy is shown in Fig. 1. In the process of annotation, annotators can back off to the more coarse-grained levels when they have low confidence on the more refined senses. This is beneficial for achieving inter-annotator reliability, especially if agreement among annotators is measured in terms of a *weighted kappa* statistic (Geertzen and Bunt, 2006), which takes into account that a tag $T_1$ at one level and a tag $T_2$ at a lower level, such that $T_2$ is dominated by $T_1$, correspond to interpretations which are not identical and hence not fully in agreement, but which are in partial agreement. Annotations could also be carried out with just the *class* level or the *class* and *type* levels while ignoring the lower level senses.

The examples in (3) illustrate the use of sense tags in the PDTB to define a specific discourse relation. Sense tags are shown in parentheses, with the colon used to illustrate the hierarchical organized sense label when the most refined subtype sense was chosen (CLASS:TYPE:SUBTYPE).

(3) a. *Big buyers like P&G say there are other spots on the globe, and in India, where the seed could be grown ...* But **no one as made a serious effort to transplant the crop.** (Comparison:Concession:Contra-expectation)

b. *Some have raised their cash positions to record levels.* Implicit=because **High cash positions help buffer a fund when the market falls.** (Contingency:Cause:Reason)

c. *But a strong level of investor withdrawal is much more unlikely this time around,* `fund managers said`. **A major reason is that investors already have sharply scaled back their purchases of stock funds since Black Monday.** (Contingency:Cause:Reason)

Discourse connectives can be ambiguous, for example *since* has a temporal sense in (4a) but a causal sense in (4b). In such cases, annotation simply involves choosing the intended sense. But connectives can also have multiple senses. For example, *since* in (4c) has both the temporal as well as the causal sense. To handle multiplicity, multiple sense tags per connective must be allowed. In the PDTB, up to two senses per connective are admitted.

(4) a. *The Mountain View, Calif., company has been receiving 1,000 calls a day about the product* since **it was demonstrated at a computer publishing conference several weeks ago.**

b. *It was a far safer deal for lenders* since **NWA had a healthier cash flow and more collateral on hand.**

c. *Domestic car sales have plunged 19%* since **the Big Three ended many of their programs Sept. 30.**

Multiplicity needs to be allowed for implicit relations as well. This is implemented by allowing multiple implicit connectives to be inserted for an implicit relation, with each connective expressing one of the two inferred senses.

The PDTB sense hierarchy contains 43 sense tags, which form the total set of discourse relations distinguished in the PDTB. This reflects the idea that there is a rather small core set of semantic relations that can hold between the situations described in the arguments of connectives (Kehler, 2002). However, the core set of relations corresponding to the 'class' level can be refined by adding other types and subtypes, and can be viewed as an open set of possible relations. The use of a hierarchically organized set of 43 discourse relations makes a basic difference between the PDTB and RST-style labeling of discourse relations (Mann and Thompson, 1988).

### 3.3. Attribution

In the PDTB, each discourse relation, whether expressed explicitly by a connective, explicitly by alternative means, or implicitly by adjacency, and each of its arguments is annotated for *attribution*, i.e. for the source to whom the relation or an argument are ascribed, such as the author(s) (or speaker) of the text, as in example (5a), or someone else who is quoted in the text, as in example (5b). Preliminary studies for the PDTB have indicated that a substantial proportion (34%) of the annotated discourse relations have another source than the author of the text, either for the relation or for one or both of its arguments.

(5) a. Since **the British auto maker became a takeover target last month,** *its ADRs have jumped about 70%.*

b. *"The public is buying the market* when **in reality there is plenty of grain to be shipped"**, said Bill Biedermann, Allendale Inc. director.

The PDTB annotation scheme distinguishes four properties of attributions, which are annotated as feature specifications: *source, type, scopal polarity*, and *determinacy*.

The *source* of an attribution distinguishes between (a) the writer of the text ("Wr"); (b) some specific other agent introduced in the text ("Ot"); and (c) some arbitrary agent indicated in the text through a non-specific reference ("Arb").

The *type* of an attribution encodes the nature of the relation beween the agent who is the source of a discourse relation and the arguments of the relation. The following kinds of relation are distinguished: (a) communication (annotated as "Comm") for asserted relations, typically involving verbs like *say, claim, argue, explain*; (b) propositional attitude ("PAtt") for cases where the source expresses a belief, expectation, assumption, etc.; (c) factive ("Ftv") for cases where the source has indicated a relation to a certain fact, e.g. by using a verb like *regret, forget, remember,* or *see*; and (d) control ("Ctrl"), for a relation to an eventuality as expressed by a control verb like *persuade, permit, promise, want*, etc.

The *scopal polarity* of an attribution serves to identify cases where verbs of attribution are negated on the surface, but where the negation in fact reverses the polarity of the attributed relation or argument, as in example (6):

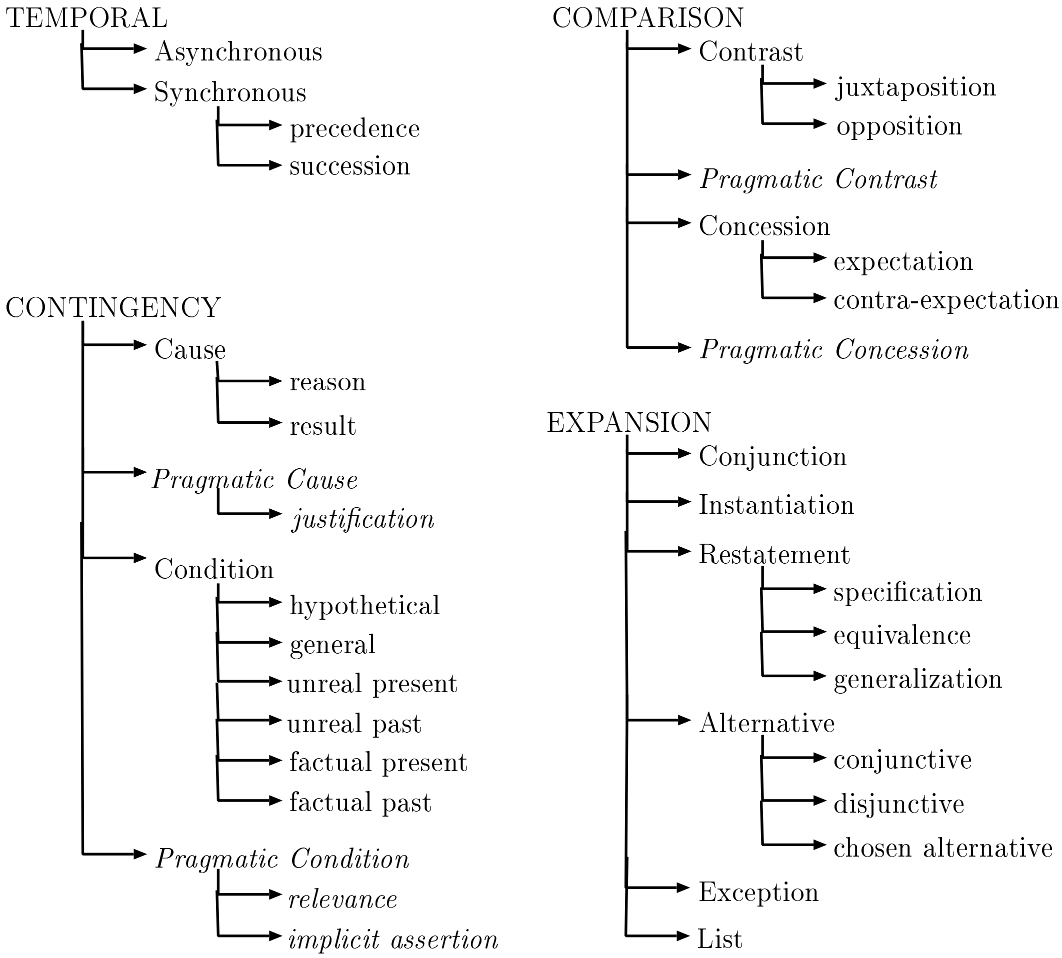


Figure 1: Hierarchy of discourse relations in the PDTB ('sense tags')

(6) *"Having the dividend increases is a supportive element in the market outlook,* but I don't think **it's a main consideration"**, he says.

In such cases, the relation (or the argument, as the case may be) is marked as having scopal polarity "Neg". This may occur both for explicit discourse relations expressed by a connective and for implicit relations.
The *determinacy* of an attribution is used to capture that the attribution may be cancelled or made indeterminate ("Ind") within a particular context, such as within the scope of a conditional or an infinitival, as in example (7), where the idea that "our teachers would educate our children better if only they got a few thousand dollars more" is not a belief or an opinion that is attributed to anyone.

(7) Its is silly libel on our teachers to think *they would educate our children better* if only **they got a few thousand dollars a year more.**

If there is no indeterminacy associated with an attribution, its determinacy has the default value "Null".

### 3.4. Representation format

In line with ISO requirements, PDTB annotations are in stand-off format: files containing the annotations are physically separate from the source text files. The PDTB annotation scheme and representation are fully described in the manual (PDTB-Group, 2008).
The current scheme for annotating a discourse relation entity in the PDTB includes a list of values, which may also represent text spans, as references to the character offsets in the source text file, and the PTB alignments of the text spans, as gorn address references to nodes in their corresponding PTB constituency trees. This may have to be revised in order to be ISO-compliant, following the joint ISO-TEI standard 24610-1 (see ISO 24610-1, 2006).

## 4. DRelML: Discourse Relations Markup Language

### 4.1. Overview

The Discourse Relations Markup Language DRelML has been designed in accordance with the ISO Linguistic Annotation Framework (LAF, ISO 24612:2009), which draws a distinction between the concepts of *annotation* and *representation*. The term 'annotation' refers to the linguistic information that is added to regions of primary data, independent of the format in which the information is represented; 'representation' refers to the format in which an annotation is rendered, independent of its content. According to LAF, *annotations* are the proper level of standardization, rather than representations. Conforming to the annotation-representation distinction, the DRelML specification follows the methodology for designing annotation languages developed in Bunt (2010), which has become standard practice in ISO work on semantic annotation. According to this methodology, the definition of an annotation language consists of three parts:

1. an abstract syntax, which specifies a class of annotation structures;
2. a formal semantics, describing the meaning of the annotation structures defined by the abstract syntax;
3. a concrete syntax, specifying a reference format for the physical representation of annotation structures defined by the abstract syntax.

Abstract and concrete syntax should moreover be related through the requirements that the concrete syntax is *complete* and *unambiguous* relative to the abstract syntax. These notions are defined as follows:

(8) a. **Completeness**: The concrete syntax defines a representation for every structure defined by the abstract syntax. (Possibly more than one, allowing alternative representations of the same abstract structure.)

b. **Unambiguity**: Every expression defined by the concrete syntax represents one and only one structure defined by the abstract syntax.

The representation format defined by a concrete syntax which has these two properties is called an *ideal representation format*. The property of 'completeness' means that there is a function $R$ which to every structure $\alpha$, defined by the abstract syntax, assigns a nonempty set $R(\alpha)$ of representations defined by the concrete syntax. Conversely, the property of 'unambiguity' means that there is a function $R^{-1}$ which assigns to every expression $e$, defined by the concrete syntax, an annotation structure $R^{-1}(e)$ defined by the abstract syntax.
An important aspect of this design methodology is that the semantics of the annotation language is defined for the *abstract* syntax; given an expression $e$ defined by the concrete syntax, its meaning is that of the annotation structure $R^{-1}(e)$. This ensures that any ideal representation format is convertible through a meaning-preserving mapping to any other ideal representation format.[2] In Ide & Bunt (2010), a mapping strategy is defined to convert from an abstract syntax to a representation in GrAF format (Ide & Suderman, 2007), and is illustrated with several annotation schemes, such as TimeML, PropBank, and FrameNet.[3] In addition to allowing for discourse annotation schemes to be represented uniformly across languages, domains, and genres, this may be useful to allow for effective combination of PDTB with GrAF renderings of PropBank and other annotations that have been done on the *WSJ*, including Penn Treebank (PTB) syntactic annotations.

[2] See Bunt (2010; 2011) for formal definitions and proofs.

[3] GrAF may be considered as a pivot format into which well-formed annotation schemes may be mapped, thus guaranteeing syntactic consistency and completeness for the purposes of comparison, merging, and transduction to other formats.

Taking the PDTB annotation scheme as the starting point for defining an ISO standard for the annotation of discourse relations, the first steps in this direction are to translate the PDTB scheme into an abstract syntax form, and to specify a concrete XML syntax for representing the annotation structures. This is the subject of the next two subsections.

### 4.2. Abstract syntax

The abstract syntax of DRelML consists of: (a) a specification of the elements from which annotation structures are built up, a 'conceptual inventory', and (b) a specification of the possible ways of combining these elements.

**a. Conceptual inventory**

The conceptual inventory of DRelML consists of a number of disjoint sets whose elements provide the ingredients for building annotation structures for discourse relations. Since a discourse relation in the PDTB is always a binary relation, with two arguments, the ingredients we need are those for identifying a discourse relation and its two arguments, including their attributions.

Since annotations add linguistic information to certain regions of primary data, such as particular stretches of text or speech, the annotation of a discourse relation includes the identification of the regions of primary data corresponding to the arguments of the relation, and in the case of an explicit discourse relation (expressed by a connective or by another type of expressions) also the region where the relation is expressed. In stand-off format, this is done through pointers to the primary data or to elements at another layer of annotation where the regions of primary data are identified. Following ISO practice, we will use the term 'markable' to refer to the entities that anchor an annotation directly or indirectly in the primary data. The conceptual inventory therefore also includes a set of markables. Altogether, the conceptual inventory therefore consists of the following sets:

1. *DR*, a finite set of discourse relations, $R_1, R_2, \ldots, R_n$. The hierarchical organization of the PDTB set of discourse relations, with lower tiers expressing more fine-grained meanings, is as such not part of the conceptual inventory, but follows from the definitions of each of these relations (cf. (Miltsakaki et al., 2008)).
2. *EntRel*, a singleton set containing a coherence relation, expressing that two sentences are related due to semantic relations between entities mentioned in the two sentences, such as coreference.
3. *MA*, a finite set of markables to which discourse relations information can be attached.
4. Four finite sets of features of attributions – *source*, *type*, *polarity*, and *determinacy*: *AtS* (attribution source), *AtT* (attribution type), *AtP* with two values for scopal polarity, and *AtD* with two values for the determinacy of an attribution.

5. *AOType*, a finite set of abstract object semantic types, $ao_1, ao_2, \ldots, ao_n$. Compared to the PDTB this is a new annotation category that we have introduced in order to make room for specifying semantic information about arguments, if desired. As with the discourse relations, inheritance relations hold between object types; these are based on the hierarchical classification in Asher (1993).

**b. Annotation structures**

An annotation structure is a set of *entity structures* and *link structures*. An entity structure contains semantic information about a region of primary data, as identified by markables; a link structure describes a semantic relation between the contents of two such regions. DRelML annotations can refer to six kinds of markables, described below.

**Entity structures:** An entity structure is one of the following structures:

a. *Explicit Attribution Entity Structure*, which is a pair ⟨*m, a*⟩ consisting of a markable $m$ and an an *'Attribution Information Structure'* $a$, which is one of the following structures:

- ⟨*as*⟩;
- ⟨*as, at*⟩;
- ⟨*as, ap, ad*⟩;
- ⟨*as, at, ap, ad*⟩,

where $m \in MA$, *as* ∈ *AtS*, *at* ∈ *AtT*, *ap* ∈ *AtP*, and *ad* ∈ *AtD*,

The different possible structures capture the fact that, if attribution is annotated for discourse relations and their arguments, the scheme is still flexible with respect to what exactly is annotated. Minimally, only the text span signaling the attribution is marked and a source. In the other structures, one or more additional semantic features are also annotated, including the semantic type, polarity and determinacy of the attribution.

As the name suggests, Explicit Attribution Entity Structures will be used to annotate explicit attributions, while Attribution Information Structures will be used for annotating implicit ones. For short, we will also use the term *Attribution Structure* to designate either an Explicit Attribution Entity Structure or an Attribution Information Structure.

b. *Explicit Relation Entity Structure*, which is one of the following structures:

1. ⟨*m, r*⟩; ⟨*m, r, a*⟩; ⟨*m, r,*, $m_{hd}$, $m_{mod}$⟩; ⟨*m, r, a*, $m_{hd}$, $m_{mod}$⟩;
2. ⟨*m*, $r_1$, $r_2$⟩; ⟨*m*, $r_1$, $r_2$, *a*⟩; ⟨*m*, $r_1$, $r_2$, $m_{hd}$, $m_{mod}$⟩; ⟨*m*, $r_1$, $r_2$, *a*, $m_{hd}$, $m_{mod}$⟩.

where *m* is a markable, *r*, $r_1$, $r_2$ ∈ *DR* are discourse relations, *a* is an Attribution Structure, and $m_{hd}$ and $m_{mod}$ are markables identifying the head and modifier(s) of a discourse connective, respectively.

The phenomenon that discourse connectives can have multiple senses is captured by the possible structures in (ii), with two senses ($r_1$ and $r_2$). Only up to two senses are allowed. Note that all structures occur with and without an Attribution Structure and with and without a connective head and modifier specification. This means that these elements are optional.

c. *Argument Entity Structure*, which is one of the following structures:

⟨*m*⟩; ⟨*m, a*⟩; ⟨*m, a, ao*⟩

where *m* is a markable, *a* is an *Attribution Structure*, and *ao* ∈ *AOType* is an abstract object type. Three different structures are defined, in order to allow the argument to be annotated with an attribution and/or with an abstract object type, without making any of them obligatory.

**Link structures:** A link structure is one of the following:

- An *Explicit Discourse Relation Structure*, which is a triple ⟨*Arg1, Arg2, R*⟩, consisting of two *Argument Entity Structures*, *Arg1* and *Arg2*, and an *Explicit Relation Entity Structure*, *R*.
- An *Implicit Discourse Relation Structure* is one of the following structures:

  i. ⟨*Arg1, Arg2, r*⟩; ⟨*Arg1, Arg2, r, a*⟩,

  ii. ⟨*Arg1, Arg2*, $r_1$, $r_2$⟩; ⟨*Arg1, Arg2*, $r_1$, $r_2$, *a*⟩

  where *Arg1* and *Arg2* are *Argument Entity Structures*, *r*, $r_1$,$r_2$ ∈ *DR* are discourse relations, and *a* is an Attribution Structure.

  As in the case of an Explicit Relation Entity Structure, the two variants in ii. capture the phenomenon that two sentences may be semantically related by more than one discourse relation (maximally two); the occurrence of variants with and without an Attribution Structure means that attributions of arguments are treated as optional.
- An *Entity Relation Structure*, ⟨*Arg1, Arg2, E*⟩ consisting of the entity-based coherence relation $E$t and two arguments *Arg1, Arg2*, which are either just a markable ⟨*m*⟩ or a pair ⟨*m, a*⟩ where *ao*∈*AOType* is an abstract object type.

### 4.3. Concrete syntax

Given the abstract syntax defined above, an XML-based concrete syntax of DRelML is defined by applying the notion of an ideal representation format, defined above. As described in Bunt (2010), an ideal XML-based representation format can be defined systematically by designing XML elements and attributes to correspond to object types and their properties. For DRelML this means the definition of the following representation structures.

1. For each type of entity structure, defined by the abstract syntax, define an XML element with the following attributes:

(a) one for each component of the entity structure;

(b) the attribute `xml:id`, whose value is a unique identifier of the entity structure;

(c) the attribute `target`, whose value refers to a markable.

2. For each type of link structure, define an XML element with attributes whose values represent a relation and its arguments.

The notion of an ideal representation forma allows the introduction of extra attributes and values in the concrete syntax, because of their convenience for annotators, or their usefulness for certain annotation purposes, as long as these additional components do not interfere with the requirements of completeness and umanbiguity.

Concretely, in order to be maximally compatible with the PDTB, attributes/values are introduced for representing supplementary argument regions, inserted connectives for implicit discourse relations, and the distinction between explicit discourse relations expressed by connectives and those expressed by other means ('AltLex'). Altogether, this leads to the following concrete syntax definition:

**Entity structure representations**

1. an XML element called `dRelArgument`, wich has the following attributes:
   - `xml:id`, whose value specifies a unique identifier;
   - `target`, whose value identifies a markable;
   - `attribution`, whose value represents an explicit or implicit attribution (*optional*);
   - `aoType`, whose value specifies the abstract object type denoted by the markable (*optional*);
   - `supplRegion`, whose value represents a supplementary markable (*optional*).
2. an XML element called `explDRel`, which has the following attributes:
   - `xml:id`, whose value specifies a unique identifier;
   - `target`, whose value represents a relational markable;
   - `synType`, whose value indicates whether an explicit discourse relation is expressed by a connective (the value `connective`) or by some other kind of expression (the value `altLex`) (*optional*);
   - `headConn`, whose value represents the lexical head of a discourse relation expressed by a connective (*optional*);
   - `modConn`, whose value represents the modifier, if present, of a discourse relation expressed by a connective (*optional*);
   - `attribution`, whose value represents an explicit or implicit attribution (*optional*);
   - `discRel`, whose value names a discourse relation.
3. an XML element called `implDRel`, which has the following attributes:
   - `xml:id`, whose value specifies a unique identifier;
   - `discRel`, whose value names a discourse relation;
   - `disConn`, whose value represents a connective, inserted for an implicit discourse relation (*optional*).
4. An XML element called `explAttribution`, which has the following attributes:
   - `xml:id`, whose value specifies a unique identifier;
   - `target` whose value identifies a markable;
   - `atSource`, whose value represents the agent or other kind of source to whom a discourse relation or an argument of a relation is attributed;
   - `atType`, whose value represents the kind of attribution (*optional*; for the PDTB, the possible values are *PAtt, Ftv, Ctrl, Undef*));
   - `atPolarity`, whose value represents the scopal polarity, possibly associated with a negated discourse relation (*optional*);
   - `atDeterminacy`, whose value represents the determinacy of the attribution (*optional*).
5. An XML element called `implAttribution`, which has the same attributes as an `explAttribution`, except that it does not have a `target` attribute, being a non-consuming tag.

**Link structure representations**

- an element called `discourseRelation`, which has the following attributes:
  - `xml:id`, whose value specifies a unique identifier;
  - `arg1` and `arg2`, whose values are `dRelArgument` elements representing the arguments of the relation;
  - `rel1` and `rel2`, whose values are both either an `explDRel` or an `implDRel` element, representing the explicit or implicit discourse relations between the two arguments; `rel1` is obligatory; `rel2` is optional and used only when the two arguments are related by two discourse relations.
- an element called `entityRelation` which has two attributes: `arg1` and `arg2`, whose values refer to two `dRelArgument` elements, and the attribute `rel` which has the value `entityRel`;

## 5. Examples

(9) Example of the representation of a simple explicit discourse relation, with temporal connective *since*:

```
<dRelML>
<discourseRelation xml:id="dr1"
  arg1="#a1"
  arg2="#a2"
  rel="#er1"/>
<dRelArgument xml:id="a1"
  target="#m1"
  attribution="#at1"/>
<dRelArgument xml:id="a2"
  target="#m3"
  attribution="#at1"/>
<explRel xml:id="er1"
  target="#m2"
  discRel="succession"
  attribution="#at1"/>
<attributionInfo xml:id="at1"
  aSource="ot"/>
</dRelML>
```

(10) Example of the representation of a multifunctional discourse marker, with the connective *since* in temporal and causal interpretation:

```
<dRelML>
<discourseRelation xml:id="dr1"
  arg1="#a1"
  arg2="#a2"
  rel1="#er1"
  rel2="#er2"/>
<dRelArgument xml:id="a1"
  target="#m1"
  attribution="#at1"/>
<dRelArgument xml:id="a2"
  target="#m3"
  attribution="#at1"/>
<explRel xml:id="er1"
  target="#m2"
  discRel="succession"
  attribution="#at1"/>
<explRel xml:id="r2"
  target="#m2"
  discRel="reason"
  attribution="#at1"/>
<implAttribution xml:id="at1"
  aSource="ot"/>
</dRelML>
```

(11) An implicit simple discourse relation (conjunction), with different attribution sources of the two arguments:

```
<dRelML>
<discourseRelation xml:id="dr1"
  arg1="#a1"
  arg2="#a2"
  rel="#ir1"/>
<dRelArgument xml:id="a1"
  target="#m1"
  attribution="#at1"/>
<dRelArgument xml:id="a2"
  target="#m2"
  attribution="#at2"/>
<explAttribution xml:id="at1"
  target="#m3"
  aSource="ot"
  aType="comm"/>
<implAttribution xml:id="at2"
  aSource="wr"/>
<implRel xml:id="ir1"
  discRel="conjunction"
  attribution="#at1"/>
</dRelML>
```

(12) An implicit multiple discourse relation (conjunction and comparison):

```
<dRelML>
<discourseRelation xml:id="dr1"
  arg1="#a1"
  arg2="#a2"
  rel1="#ir1"
  rel2="#ir2"/>
<dRelArgument xml:id="a1"
  target="#m1"
  attribution="#at1"/>
<dRelArgument xml:id="a2"
  target="#m2"
  attribution="#at2"/>
<attributionInfo xml:id="at1"
  target="#m3"
  aSource="ot"
  aType="comm"/>
<attributionInfo xml:id="at2"
  aSource="wr"/>
<implRel xml:id="ir1"
  discRel="conjunction"
  attribution="#at1"/>
<implRel xml:id="ir2"
  discRel="comparison"
  attribution="#at1"/>
</dRelML>
```

## 6. Conclusions and perspectives

The exercise of creating an abstract syntax for the PDTB annotation scheme and rendering it in a graphic form shows the structure of the annotations clearly. The resulting concrete syntax is much more readable than the original format, and therefore errors and inconsistencies may be more readily identified. Furthermore, because it is rendered in XML, annotations can be validated against an XML schema (including validation that attribute values are among a list of allowable alternatives).

The abstract syntax also shows clearly an overall structure for a general-purpose standard for annotating discourse relations. We envision that any general-purpose discourse annotation scheme must allow for annotation based on all or any of several perspectives on elements of the task, such as semantic, interpersonal/intentional, and stylistic/textual, as identified in Hovy (1995). PDTB annotations are classified as "informational" (semantic, inter-propositional, ideational, pragmatic); the intentional and textual perspectives lie outside the scope of PDTB. PDTB's attribution types and the set of semantic classes, combined with those of other schemes, provide a base for a structured set of discourse annotation classes for the ISO specification along the various axes of perspective, and at different levels of granularity.

Several topics for further work in developing an ISO standard for discourse relation annotation have emerged during the work reported in this paper. First, the approach underlying the PDTB has limited its scope to the annotation of relations between adjacent sentences. This limitation has been motivated by practical considerations regarding the work of human annotators. From a semantic point of view, however, both discourse relations within sentences and between non-adjacent sentences may be important. Second, the formal semantics of the abstract syntax still has to be worked out. Third, the establishment of sets of annotation concepts that are more broadly important than for the WSJ should deserve careful consideration, taking a range of languages, domains, and genres into account. This concerns in particular the set of discourse relations, and the sets of values used for the characterization of attributions (such as the set *Writer, Other, Arbitrary, Inherited* used in the PDTB). Explicit definitions of all the concepts, finally chosen as part of the standard, will have to be provided, and inserted in the ISOCat data registry.[4] Finally, the standard will not only have to define annotation and representation structures and concepts, but also examples and guidelines for their use in a range of practical situations.

## 7. References


Nicholas Asher (1993) *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.

Nicholas Asher and Alex Lascarides (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Harry Bunt (2010) A methodology for defining semantic annotation languages exploiting syntactic-semantic isomporphisms. In: Alex Chengyu Fang, Nancy Ide and Jonathan Webster (eds.) *Proceedings of ICGL 2010, Second International Conference on Global Interoperability for Language Resources*, Hong Kong, pp. 29-45.

Harry Bunt (2011) Defining languages for semantic annotation with an abstract syntax and a formal semantics. *Journal of Natural Language Engineering*, to appear.

Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum (2010) Towards an ISO standard for dialogue act annotation. In *Proceedings of LREC 2010*.

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, and David Traum (2010) ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of LREC 2012*.

Jeroen Geertzen and Harry Bunt (2006) Measuring annotator agreement in a complex hierrchical dialogue act scheme. in *Proceedings of the $7^{th}$ SIGdial Workshop on Discourse and Dialogue*, Sidney, pp. 126-133.


[4]See e.g. Kemps-Snijders, Windhouwer and Wright (2010) and `http://www.isocat.org`.


Barbara J. Grosz and Candace L. Sidner (1986) Attention, intentions, and the structure of discourse. In *Computational Linguistics* Vol. 12, n.3. pages 175-204.

Jerry Hobbs (1985) *On the coherence and structure of discourse*. Technical Report. Stanford University.

Ed Hovy (1995) The Multifuncitonality of Discourse Markers. in *Proceedings of the Workshop on Discourse Markers*, Egmond-aan-Zee, The Netherlands.

Nancy Ide and Harry Bunt (2010). Anatomy of Annotation Schemes: Mappings to GrAF. in *Proceedings of LAW-IV: the Fourth Linguistic Annotation Workshop*, Uppsala, pp. 115 -124.

Nancy Ide, Rashmi Prasad, and Aravind Joshi (2011). Towards Interoperability for the Penn Discourse Treebank, In *Proceedings of the 6th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-6)*, Oxford, U.K.

Nancy Ide and Laurent Romary (2004). International Standard for a Linguistic Annotation Framework. *Natural Language Engineering*, 10:211 - 225.

Nancy Ide and Keith Suderman (2007). GrAF: A Graph-based Format for Linguistic Annotations. In *Proceedings of the First Linguistic Annotation Workshop LAW-I*, held in conjunction with ACL 2007. Prague, pp. 1-8.

ISO (2006) ISO 24612:2006 Language resource management: Feature structures, Part 1: Feature structure representation. ISO, Geneva.

ISO (2010) ISO 24612:2010 Language resource management: Linguistic annotation framework (LAF), ISO, Geneva.

ISO (2010) ISO DIS 24617-2 Language resource management – Semantic annotation framework – Part 2: Dialogue acts. ISO, Geneva.

Andrew Kehler (2002) *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford.

Marc Kemps-Snijders, Menzo Windhouwer, and Sue Ellen Wright (2010) Standardizing Data Cetegories in ISOcat: Implementing Group Work for Thematic Domains. In *Proceedings Post-conference workshop at TKE 2010 Conference on Terminology and Knowledge Engineering*, Dublin.

Alan Lee, Rashmi Prasad, Aravind Joshi and Bonnie Webber (2008) Departures from tree structures in discourse: Shared arguments in the Penn Discourse Treebank. In *Proceedings of the Constraints in Discourse III Workshop*. Potsdam, Germany.

William Mann and Sandra Thompson (1988). Rhetorical structure theory. Toward a functional theory of text organization. *Text* 8(3):243-281.

Mitchell P. Marcus, Beatrice Santorini and Mary Ann Marcinkiewicz (1993) Building a large annotated corpus of English: The Penn Treebank. In *Computational Linguistics* Vol.19, n.2. pages 313-330.

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi and Bonnie Webber (2004) Annotating discourse connectives and theoir arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*. Boston, MA. Pages 9-16.

Eleni Miltsakaki, Livio Robaldi, Alan Lee and Aravind Joshi (2008) Sense Annotation in the Penn Discourse

Treebank. In *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science* Vol. 4919. Pages 275-286.

Livia Polanyi (1987) *The Linguistic Discourse Model: Towards a formal theory of discourse structure*. Technical Report. Bolt Beranek and Newman, Inc.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Aravind Joshi and Bonnie Webber (2007). Attribution and its Annotation in the Penn Discourse TreeBank. In *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse* Vol.47, n.2. pages 43-64.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

The PDTB Group (2008) *The Penn Discourse Treebank 2.0. Annotation Manual*. IRCS Technical Report IRCS-08-0.1 Institute for Research in Cognitive Science, University of Pennsylvania. Philadelphia, PA.

The Text Encoding Initiative (2007) P5 Guidelines for Electronic Text Encoding and Interchange, edited by Syd Bauman and Lou Burnard. The Text Encoding Initiative, Charlottesville, Virginia. Available at `http://www.tei-c.org/Guidelines/P5/`.

Bonnie Webber, Aravind Joshi, Matthew Stone, and Alistair Knott (2003). In *Computational Linguistics* Vol.29, n.4. pages 545-587.

Florian Wolf and Edward Gibson (2005) Representing discourse coherence: A corpus-based study. In *Computational Linguistics* Vol.31, n.2.

# The Current Status of ISO-Space

**James Pustejovsky, Jessica L. Moszkowicz, Marc Verhagen**

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

**Abstract**

We report on ISO-Space version 1.4, an annotation specification for capturing spatial and spatiotemporal information in natural language that is now in its fourth incarnation. This version substantially improves upon earlier ISO-Space specifications in a few notable ways. The representation of locations is no longer overloaded such that geolocations have a more complete annotation and non-geolocations are captured with specific tags. In addition, interactions with existing annotation standards such as TimeML have been clarified. The treatment of spatial prepositions has been modified so that their annotation is more suggestive of what spatial relationships should hold between two spatial objects. Finally, spatial relationships are now captured with four distinct link tags: qualitative spatial links for topological relationships, orientation links for non-topological relations, movement links for motion, and measure links for detailing a metric relationship between two spatial objects or what the dimensions of a particular object are. The most recent version of the specification is presented with illustrative examples. We conclude with some outstanding issues that have yet to be captured in the specification.

## 1. Introduction

This document presents the current specification of ISO-Space, an emerging standard for the annotation of spatial and spatiotemporal information.[1] The goal of this specification is to provide the syntax of ISO-Space through descriptions of its tags and their attributes and examples that illustrate their basic use. The ISO-Space Annotation Guidelines will provide a fuller discussion of how to use this specification. ISO-Space incorporates the annotations of static spatial information, borrowing from the SpatialML scheme (Mani et al., 2010; Mani et al., 2008), and events, borrowing from the TimeML scheme (Pustejovsky et al., 2005).

ISO-Space is being developed as a Work Item within the ISO/TC37/SC4. The core working group includes, besides the authors: Harry Bunt, Kiyong Lee, Inderjeet Mani, and Annie Zaenen. It assumes the ISO CD 24612 Language Resource Management - Linguistic Annotation Framework standard (Ide and Romary, 2004). As such, ISO-Space provides a stand-off annotation scheme with some tags in the specification linked explicitly to text offsets and others representing relationships between other tags.

We begin with the basic tags of ISO-Space: the tags that include text offsets (in most cases) and describe the basic spatial elements within a document. We then turn to the link tags, which capture more complex spatial information by relating the basic elements together.

## 2. Basic ISO-Space Elements

### 2.1. Location Tags

Locations in ISO-Space come in two varieties: PLACE and PATH. Each of these tags captures a specific kind of spatial information in the text and they can both subsequently participate in spatial relationships by way of the link tags.

PLACE **Tag.** The PLACE tag is inherited from SpatialML. This tag is used to annotate geographic entities likes lakes and rivers, as well as administrative entities like towns and counties. As in SpatialML, a PLACE tag in ISO-Space must be directly linked to an explicit span of text. Some examples of this tag are presented in (1). Note that additional spatially relevant elements in these sen-

[1] ISO-Space was first developed following a meeting at Brandeis University in 2009 and then refined at two workshops in 2010 and 2011. For a description of version 1.3, see (Pustejovsky et al., 2011).

tences are left unmarked for now; only the PLACEs are shown.

(1) a. A Libyan ship that tried to break Israel's blockade of [**Gaza**$_{pl1}$] docked in the [**Egyptian port of El Arish**$_{pl2}$] on Thursday afternoon.

b. The new tropical depression was about 430 miles (690 kilometers) west of the [**southernmost Cape Verde Island**$_{pl3}$], forecasters said.

The attributes for the PLACE tag are largely inherited from SpatialML.[2] For example, for those places that have known latitude and longitude values, the `latLong` attribute can be used to allow for mapping to other resources such as Google Maps. ISO-Space also includes the *Document Creation Location* or `DCL` attribute. This is a special place that serves as the "narrative location". If the document includes a DCL, it is generally specified at the beginning of the text, similarly to the manner in which a Document Creation Time (DCT) is specified in TimeML. If a place is the DCL, this is marked with a special attribute in the annotation of the place. The current set of PLACE attributes is shown in Figure 1.

The values for the `type` attribute are identical to the values from the SpatialML PLACE tag with the exception of some types such as VEHICLE, which is a spatial named entity in ISO-Space, and ROAD, which is a path. Places can be in the form of proper names (*New York*) or nominals (*town*), which are marked with the `form` attribute as NAM or NOM, respectively. Examples of what constitutes each `type` can be found in complete ISO-Space Specification, available at `www.iso-space.org`.

The place's `mod` attribute is there to capture cases like *tall building*, *the higher observation deck* and *two towers*, where *tall*, *higher* and *two* do not really constrain the location of the entity but they do add spatial information.

The `mod` attribute is substantially different from its counterpart in SpatialML where it was used for modifiers like *bottom of the well*, *Burmese border*, *near Harvard*, *northern India* and *the right side of the building*. In many cases, these modifiers were deemed necessary in SpatialML because it focuses on annotating gazetteer entries. In ISO-Space, these cases are analyzed in two ways: (i) the SpatialML modifier is a signal for spatial relations or (ii) the entire phrase is a place.[3]

**PATH Tag.** A PATH is a location where the focus is on the potential for traversal or functions as a boundary. This includes common nouns like *road* and *river* and proper names like *Route 66* and *Kangamangus Highway*.

Paths typically have begin points and end points, although these are often not expressed in the text. Example (2) shows an instance of a PATH in which the endpoints happen to be explicit.

(2) the [**rail road**$_{p1}$] from [**Boston**$_{pl1}$] to [**Maine**$_{pl2}$]

Path(p1, beginPoint=pl1, endPoint=pl2)

The attributes of PATH are a subset of the attributes of the PLACE element, but with the `beginID` and `endID` elements added as shown in Figure 2.

### 2.2. Non-Location Tags

Location tags essentially designate a region of space that can be related to other regions of space. In addition, ISO-Space allows for non-location elements of a text to be coerced into behaving like a region of space so that they may participate in the same kinds of relationships. There are three of these kinds of tag in ISO-Space: SPATIAL_NE, EVENT, and MOTION. Note that, for the most part, annotating these tags should not be the responsibility of the ISO-Space annotator. Instead, capturing this kind of information should be left to other annotation schemes and it will be left to the ISO-Space annotator to recognize when such an element should participate in an ISO-Space link tag or if additional information specific to spatial language needs to be added to the annotation. Details on this will be available in the ISO-Space annotation guidelines.

[2]In fact, given a SpatialML annotation, an ISO-Space annotation should simply be able to inherit the captured PLACE elements.

[3]Given this discrepancy with SpatialML, it is likely that the ISO-Space annotator will have to perform some "clean-up" of the PLACE elements that are inherited from a SpatialML annotation. This issue will be taken up in the annotation guidelines, though, as it is not relevant to this specification.

| | |
|---|---|
| `id` | pl1, pl2, pl3, ... |
| `type` | BODYOFWATER, CELESTIAL, CIVIL, CONTINENT, COUNTRY, GRID, LATLONG, MTN, MTS, POSTALCODE, POSTBOX, PPL, PPLA, PPLC, RGN, ROAD, STATE, UTM |
| `form` | NAM or NOM |
| `continent` | AF, AN, AI, AU, EU, GO, LA, NA, PA, SA |
| `country` | a two letter ISO 3166 country code<br>See `http://www.iso.org/iso/country_codes/iso_3166_code_lists/`. |
| `state` | a principal subdivision of a country like state, province or parish, again following ISO 3661. |
| `county` | a subdivision below the state level |
| `ctv` | CITY, TOWN or VILLAGE |
| `gazref` | gazetteer name plus a colon plus an identifier, e.g. IGDB:2104656 |
| `latLong` | a coordinate from the gazetteer |
| `mod` | a spatially relevant modifier |
| `dcl` | true or false |

Figure 1: Attributes for PLACE Tag

| | |
|---|---|
| `id` | p1, p2, p3, ... |
| `beginID` | identifier of a location tag |
| `endID` | identifier of a location tag |
| `midIDs` | list of midpoint locations, if specified |
| `form` | NAM or NOM |
| `gazref` | gazetteer name plus a colon plus an identifier, e.g. IGDB:2104656 |
| `latLong` | a coordinate from the gazetteer |
| `mod` | a spatially relevant modifier |

Figure 2: Attributes for PATH Tag

**Spatial Named Entities.** A SPATIAL_NE is a named entity that is both located in space AND participates in an ISO-Space link tag. The example in (3) shows which named entities in the text are considered SPATIAL_NE tags.

(3) The new [**tropical depression**$_{sne3}$] was about 430 miles (690 kilometers) west of the southernmost Cape Verde Island, forecasters said.

When a named entity is identified as a SPATIAL_NE, it receives an `id` attribute for the ISO-Space annotation[4] and the annotator may add additional attributes as suggested in Figure 3.[5]

**Non-Motion** EVENT **Tag.** An EVENT is a TimeML event that does not involve a change of location but is directly related to another ISO-Space element by way of a link tag. Events are inherited directly from a TimeML annotation and require no further specification in ISO-Space.

MOTION **Tag.** A MOTION is a TimeML event that involves a change of location. Since motions are inherently spatial, they play a special role in ISO-Space. When a TimeML event has been identified as a MOTION, it gets re-annotated with the attributes given in Figure 4.

The `motion_type` attribute refers to the two distinct strategies for expressing concepts of motion in language: *path constructions* and *manner-of-motion constructions* (Talmy, 1985). This is illustrated in the sentences in (4), where $m$ indi-

[4]In lieu of being assigned an ISO-Space `id` number, the subsequent link tag can use the ID that was previously assigned to the entity by the layered annotation scheme.

[5]The decision to do this is likely task-based. The annotation guidelines for a specific task will instruct the annotator on what attributes to add, if any.

| | |
|---|---|
| `id` | sne1, sne2, sne3, ... |
| `form` | NAM or NOM |
| `latLong` | a coordinate |
| `mod` | a spatially relevant modifier |

Figure 3: Attributes for SPATIAL_NE Tag

| | |
|---|---|
| `id` | m1, m2, m3, ... |
| `motion_type` | MANNER or PATH |
| `motion_class` | MOVE, MOVE_EXTERNAL, MOVE_INTERNAL, LEAVE, REACH, DETACH, HIT, FOLLOW, DEVIATE, STAY |

Figure 4: Attributes for MOTION Tag

cates a manner verb, and $p$ indicates a path. In the first sentence, the motion verb specifies a path whereas in the second the motion verb specifies the manner of motion. Both are annotated as motions since the motion is implied in the manner-of-motion verb.

(4) a. John arrived$_p$ [by foot]$_m$.
b. John hopped$_m$ [out of the room]$_p$.

Motion classes are taken from (Pustejovsky and Moszkowicz, 2008), which was based on the motion classes in (Muller, 1998). These classes are associated with a spatial event structure that specifies, amongst other things, the spatial relations between the arguments of the motion verb at different phases of the event.

### 2.3. SPATIAL_SIGNAL Tag

A SPATIAL_SIGNAL is a relation word that supplies information to an ISO-Space link tag. It is typically a preposition or other function word or phrase that reveals the particular relationship between two ISO-Space elements. The attributes for this tag are given in Figure 5.
Sense information, which is stored in the `cluster` attribute is optional. The values for this attribute come from a sense inventory of spatial prepositions that is described in more detail in the annotation guidelines. The `semantic_type` attribute helps the annotator decide, along with sense information if it is available, what kind of ISO-Space relationships the signal triggers. Some examples of typical SPATIAL_SIGNALs are shown in (5).

(5) a. The book is [**on**$_{s1}$] the table.
spatial_signal(s1, cluster="on-1", semantic_type=topological, directional)

b. Boston is [**north of**$_{s2}$] New York City.
spatial_signal(s2, cluster="north_of-1", semantic_type=directional)

c. John is [**in front of**$_{s3}$] the tree.
spatial_signal(s3, cluster="in_front_of-1", semantic_type=directional)

### 2.4. MEASURE Tag

The MEASURE tag is used to capture distances and dimensions for use in a measurement link. Its attributes are shown in Figure 6. Example (6) shows the annotation of a MEASURE.

(6) The new tropical depression was about [**430 miles**$_{me1}$] ([**690 kilometers**$_{me2}$]) west of the southernmost Cape Verde Island, forecasters said.
measure(me1, value=430, unit=miles)
measure(me2, value=690, unit=kilometers)

## 3. ISO-Space Relationship Tags

There are four relation tags in ISO-Space:

(7) a. QSLINK – this represents a qualitative spatial relationship between two locations;
b. OLINK – this expresses the orientation of

| | |
|---|---|
| `id` | s1, s2, s3, ... |
| `cluster` | identifies the sense of the preposition |
| `semantic_type` | DIRECTIONAL, TOPOLOGICAL |

Figure 5: Attributes for SPATIAL_SIGNAL Tag

| | |
|---|---|
| `id` | me1, me2, me3, ... |
| `value` | number component |
| `unit` | measurement phrase component |

Figure 6: Attributes for MEASURE Tag

an location or object relative to another;
c. MOVELINK – this denotes the path of an object in motion;
d. MLINK – this defines the distance between two regions or the dimensions of a region.

Each of these tags is triggered by a specific kind of spatial element that was annotated earlier in the text. QSLINKS are introduced by topological SPATIAL_SIGNALS, OLINKS by directional SPATIAL_SIGNALS, MOVELINK by MOTION events, and MLINK by MEASURE tags.

### 3.1. Qualitative Spatial Link: QSLINK

QSLINK is used in ISO-Space to capture topological relationships between captured elements in the annotation. The attributes of QSLINK are shown in Figure 7.

The `relType` attribute values come from a slightly extended set of RCC8 relations that was first used by SpatialML. The possible values include but are not limited to DC (disconnected), EC (external connection), and IN (disjunction of tangential and non-tangential proper part).

It is worth noting that while QSLINK is used exclusively for capturing topological relationships, which are only possible between two regions, the `figure` and `ground` attributes can accept IDs for both places and paths, which are more traditional regions, as well as spatial entities and events. In the latter cases, it is actually the region of space that is associated with the location of the entity or event that participates in the QSLINK. That is, the entity or event is *coerced* to a region for the purposes of interpreting this link.

In practice, a SPATIAL_SIGNAL with a `semantic_type` of topological introduces a QSLINK as shown in example (8).

(8) a. [**The book**$_{sne1}$] is [**on**$_{s1}$] [**the table**$_{sne2}$].
spatial_signal(s1, cluster=”on-1”,
semantic_type=topological, directional)
qslink(qsl1, figure=sne1, ground=sne2,
trigger=s1, relType=EC)

b. [**The light switch**$_{sne3}$] is [**on**$_{s2}$] [**the wall**$_{sne4}$].
spatial_signal(s1, cluster=”on-2”,
semantic_type=topological, directional)
qslink(qsl2, figure=sne3, ground=sne4,
trigger=s2, relType=PO)

### 3.2. Orientation Link: OLINK

Orientation links describe non-topological relationships between spatial objects. A SPATIAL_SIGNAL with a directional `semantic_type` triggers such a link. Rather than a simple relationship type, the OLINK is built around a specific frame of reference type and a reference point. Figure 8 details the attributes for this link.

The `referencePt` value depends on the `frame_type` of the link. Absolute OLINKS have a cardinal direction as a reference point. For intrinsic OLINKS, the reference point is the same identifier that was given in the `ground` attribute. For relative OLINKS, the ID for the viewer should be provided as the reference point. If the viewer is not explicit in the text, the special value

| | |
|---|---|
| `id` | qsl1, qsl2, qsl3, ... |
| `relType` | {RCC8+} |
| `figure` | identifier of the place, path, spatial named entity, or event that is being related |
| `ground` | identifier of the place, path, spatial named entity, or event that is being related to |
| `trigger` | identifier of the spatial signal that triggered the link |

Figure 7: Attributes for QSLINK Tag

| | |
|---|---|
| `id` | ol1, ol2, ol3, ... |
| `relType` | NEAR, ABOVE, BELOW, FRONT, BEHIND, LEFT, RIGHT, NEXT TO, NORTH, ... |
| `figure` | identifier of the place, path, spatial named entity, or event that is being related |
| `ground` | identifier of the place, path, spatial named entity, or event that is being related to |
| `trigger` | identifier of the spatial signal that triggered the link |
| `frame_type` | ABSOLUTE, INTRINSIC, RELATIVE |
| `referencePt` | cardinal direction, ground entity, viewer entity |
| `projective` | TRUE, FALSE |

Figure 8: Attributes for OLINK Tag

”VIEWER” should be used. The `projective` attribute is a toggle that says whether the OLINK should have a projective interpretation. This information generally depends on what spatial signal triggered the OLINK. The examples in (9) show both projective and non-projective cases. Only the orientation links are shown.

(9) a. [**Boston**$_{pl1}$] is [**north of**$_{s1}$] [**New York City**$_{pl2}$].

olink(ol1, figure=pl1, ground=pl2, trigger=s1, relType=”NORTH”, frame_type=ABSOLUTE, referencePt=NORTH, projective=TRUE)

b. [**The dog**$_{sne1}$] is [**in front of**$_{s2}$] [**the couch**$_{sne2}$].

olink(ol2, figure=sne1, ground=sne2, trigger=s2, relType=”FRONT”, frame_type=INTRINSIC, referencePt=sne2, projective=FALSE)

c. [**The dog**$_{sne3}$] is [**next to**$_{s3}$] [**the tree**$_{sne4}$].

olink(ol3, figure=sne3, ground=sne4, trigger=s3, relType=”NEXT TO”, frame_type=RELATIVE, referencePt=VIEWER, projective=FALSE)

d. [**The hill**$_{pl3}$] is [**above**$_{s4}$] [**the town**$_{pl4}$].

olink(ol4, figure=pl3, ground=pl4, trigger=s4, relType=”ABOVE”, frame_type=INTRINSIC, referencePt=pl4, projective=TRUE)

e. [**The helicopter**$_{pl5}$] is [**above**$_{s5}$] [**the town**$_{pl6}$].

olink(ol4, figure=pl5, ground=pl6, trigger=s5, relType=”ABOVE”, frame_type=INTRINSIC, referencePt=pl4, projective=FALSE)

f. [**The book**$_{sne1}$] is [**on**$_{s1}$] [**the table**$_{sne2}$].

olink(ol4, figure=sne1, ground=sne2, trigger=s1, relType=”ABOVE”, frame_type=INTRINSIC, referencePt=sne2, projective=FALSE)

g. [**The light switch**$_{sne3}$] is [**on**$_{s2}$] [**the wall**$_{sne4}$].

olink(ol4, figure=sne3, ground=sne4, trigger=s2, relType=”ABOVE”, frame_type=INTRINSIC, referencePt=sne2, projective=FALSE)

### 3.3. Movement Link: MOVELINK

Movement links, which are generally introduced by motion events, capture information about the path a particular motion takes. It has the attributes shown in Figure 9.
For example:

(10) [**John**$_{sne1}$] [**walked**$_{m1}$] from [**Boston**$_{pl1}$] to [**Cambridge**$_{pl2}$].

movelink(mv1, trigger=m1, source=pl1, goal=pl2, mover=sne1, goal_reached=TRUE)

### 3.4. Metric Link: MLINK

Metric relationships are captured with the MLINK tag. This tag can either describe the metric relationship between two spatial objects or the dimensions of a single object. The attributes are given in Figure 10.
When MLINK is used to describe an internal dimension of an object, the ID of the object should appear in the `figure` attribute. The annotator may either repeat the ID in the `ground` attribute or leave this attribute out. The examples below show several ways in which MLINK is used. Examples (11c) and (11d) show the unique case when a stative path, or a path that does not involve traversal, is used to describe the dimensions of a location. In such a case, the optional attributes `endPoint1` and `endPoint2` are used.

(11) a. The new [**tropical depression**$_{sne1}$] was about [**430 miles**$_{me1}$] ([**690 kilometers**$_{me2}$]) west of the [**southernmost Cape Verde Island**$_{pl1}$], forecasters said.

mlink(ml1, relType = DISTANCE, figure=sne1, ground=pl1, val=me1)

b. [**The football field**$_{sne2}$] is [**100 yards**$_{me2}$] long.

mlink(ml2, relType = LENGTH, figure=sne2, ground=sne2, val=me2)

c. [**Times Square**$_{pl2}$] stretches from [**42nd**$_{p1}$] to [**47th streets**$_{p2}$].

mlink(ml3, relType = GENERAL_DIMENSION, figure=pl2, ground=pl2, endPoint1=p1, endPoint2=p2)

d. [**The office**$_{pl3}$] stretches for [**25 feet**$_{me3}$] from [**the bookcase**$_{sne3}$] to [**the white board**$_{sne4}$].

mlink(ml4, relType=GENERAL_DIMENSION, figure=pl4, ground=pl3, val=me3, endPoint1=sne3, endPoint2=sne4)

e. [**The hot dog stand**$_{sne5}$] near [**Macy's**$_{sne6}$].

mlink(ml5, relType=GENERAL_DIMENSION, figure=sne5, ground=sne6, val=NEAR)

## 4. Outstanding Issues

The above specification leaves several issues unanswered concerning the representation of spatial information as a specification language. Perhaps the most significant is the absence of a *native* representation of the 3D objects denoted by linguistic expressions, along with the associated functions we naturally ascribe to them. For example, the topological relationship between a glass and the liquid it holds should convey more than the RCC8 relations of EC or TPP, neither of which is exactly correct. Rather, within a 3D interpretation, the appropriate relation should express containment of a region within a convex volume. Similarly, the interpretation of an object inside a box should also make reference to such a containment relation, rather than a mere EC value. That is, more credence should be given to image-schematic accounts of spatial categories and how this impacts the spatial configurational relation that are denoted by real-world spatial situations (Frank and Raubal, 1999; Kuhn, 2007). This is currently being examined within the ISO-Space working group for inclusion into the specification language.

## Acknowledgements

We would like to thank the members of the ISO-Space Working Group for their significant input to the current specification as well as the participants of the Airlie Workshop. Part of this research was funded under the NURI grant HM1582-08-1-0018 by the National Geospatial Agency.

## 5. References

Andrew U. Frank and Martin Raubal. 1999. Formal specification of image schemata – a step towards interoperability in geographic information systems. *Spatial Cognition and Computation*, 1:67–101. 10.1023/A:1010004718073.

| | |
|---|---|
| id | mvl1, mvl2, mvl3, ... |
| trigger | identifier of the motion event that triggered the link |
| source | identifier of the place, path, spatial named entity, or event at the beginning of the path |
| goal | identifier of the place, path, spatial named entity, or event at the end of the path |
| mover | identifier of the entity that moves along the path |
| goal_reached | TRUE, FALSE |
| pathID | identifier of a path that is equivalent to the one described by the MOVELINK |

Figure 9: Attributes for MOVELINK Tag

| | |
|---|---|
| id | ml1, ml2, ml3, ... |
| figure | identifier of a spatial object |
| ground | identifier of the related spatial object, if there is one |
| relType | DISTANCE, LENGTH, WIDTH, HEIGHT, GENERAL_DIMENSION |
| val | NEAR, FAR, identifier of a measure |
| endPoint1 | identifier of a spatial object at one end of a stative path |
| endPoint2 | identifier of a spatial object at the other end of a stative path |

Figure 10: Attributes for MLINK Tag


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

W. Kuhn. 2007. An image-schematic account of spatial categories. In *Lecture Notes in Computer Science 4736*, pages 152–168.

Inderjeet Mani, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby, and Ben Wellner. 2008. Spatialml: Annotation scheme, corpora, and tools. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may. European Language Resources Association (ELRA).

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.

Philippe Muller. 1998. A qualitative theory of motion based on spatio-temporal primitives. In Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, editors, *KR'98: Principles of Knowledge Representation and Reasoning*, pages 131–141. Morgan Kaufmann, San Francisco, California.

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, Robert Ingria, Roser Sauri, Jose Castano, Jessica Littman, Robert Gaizauskas, Andrea Setzer, Graham Katz, and Inderjeet Mani. 2005. The specification language timeml. In Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas, editors, *The Language of Time*. Oxford University Press.

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.

Leonard Talmy. 1985. Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen, editor, *Language typology and semantic description Volume 3: Grammatical categories and the lexicon*, pages 36–149. Cambridge University Press.