Multi-layered Annotation of Non-textual Data for Spatial Information

Kiyong Lee
Korea University, Department of Linguistics
Seoul, Korea
ikiyong@gmail.com

Abstract

Spatial and spatio-temporal information is often carried by non-textual data such as maps, diagrams, tables, or pictures, both still and moving, either embedded in a text or standalone. The annotation of nontextual data raises the following questions: (i) what are the markables and how should they be coded? (ii) how should relevant information be inferred which is implicit in the data? We answer these questions with a multilayered approach.

1 Introduction

Non-textual data such as maps, figures, or pictures, either still or moving, are powerful media that carry spatial or spatio-temporal information. This paper concerns the annotation of such data, whether they are embedded in a text or presented alone. As its basic annotation scheme, it follows *ISO-Space*, a semantic annotation scheme which was proposed by Pustejvosky et al. (2012) for the annotation of spatial information in natural language. It is claimed that *ISO-Space* can be adequately applied to the annotation of non-textual data as well as text data in natural language.

Section 2 presents partial specification of *ISO-Space*, section 3 discusses making references to markables, section 4 deals with understanding conventions, section 5 illustrates multi-layered annotation, and section 6 makes concluding remarks.

2 Partial Specification of *ISO-Space*

Given a text (fragment) t_L of a language L, the annotation scheme $\mathcal{AS}_{isoSpace}$ of *ISO-Space* can be defined formally as a quadruple $<M, E, R, @>$, where M is a nonempty finite set of (some selected) segments of t_L, called *markables*, E a nonempty finite set of elements, called *basic entities*, which are either atomic or composite, R an n-ary (basically binary) relation over E, and $@$ a set of functions from a set of attributes to a set of values for each element e in E and each relation r in R. One particular attribute is an attribute, named `@target`, that anchors a basic atomic entity e in E to a markable m in M. For the general formulation of an annotation scheme $\mathcal{AS}$, we basically follow Lee (2012), which is slightly different from that of Bunt (2010) or Bunt (2011).

The set M of markables consists of all the expressions, i.e., sequences of tokens or words in t_L, that refer to all of the basic entities of each of the types defined by E. These entities include (1) spatial entities, tagged as `PLACE` and `PATH` or (2) entities that are not genuinely spatial, but involve spatial entities, tagged as `EVENT, MOTION, SPATIAL_NE (named entity)` or `SPATIAL_SIGNAL`. The set R of n-ary links over E include (1) qualitative spatial link, (2) orientation link, (3) movement link, and (4) metric link tagged as `QSLINK, OLINK, MOVELINK`, and `MLINK`, respectively.

The specification of sets of attribute-value pairs for each of the basic entity types and the links requires a complex listing. Each basic entity e in E and each link r in R has a unique ID, specified with the attribute `@xml:id` in XML representation. Each basic entity e is anchored to a markable in M, specified with the attribute `@target` in standoff annotation and assigned a sequence of tokens as value if t_L is a *tokenized* text. Note that there are two types of basic entities, *atomic* and *composite*. Atomic basic entities are simply anchored to a markable in M, whereas composite basic entities are anchored to other basic entities as well as to markables. The entity type `PLACE`, for instance, is an *atomic* entity type, while the entity type `PATH` is a *composite* entity type, for the latter is anchored to `PLACE`s.

Instead of presenting $\mathcal{AS}_{isoSpace}$ as a whole as is formally defined, we may introduce it only partially and also in an informal way with some illustrations. For this, consider the following text:

(1) Mia drove to Jeju International Airport yesterday.

This sentence contains 8 tokens including a punctuation mark. Out of them, *ISO-Space* selects 6 tokens and treats them as four markables, "Mia", "drove", "to", and "Jeju International Airport", as shown below:

(2) Mia$_{token1}$ drove$_{token2}$ to$_{token3}$ Jeju$_{token4}$ International$_{token5}$ Airport$_{token6}$ yesterday.

Corresponding to the four markables, four basic entities are introduced: `SPATIAL_NE, MOTION, SPATIAL_SIGNAL` and `PLACE`. A link is also introduced: `<MOVELINK>`. Each of them is specified with a list of appropriate attribute-value assignments with some modifications on the current list of *ISO-Space*, as is represented in XML as follows:[1]

(3)
```
<isoSpace>
<SPATIAL_NE xml:id="sne1"
target="#token1" type="PERSON"
form="NAME"/>
<MOTION xml:id="m1"
target="#token2"
motion_type="MANNER"
motion_class="MOVE_EXTERNAL"/>
<SPATIAL_SIGNAL xml:id="s1"
target="#token3"/>
<PLACE xml:id="pl1"
target="#(token4,token6)"
type="FAC" subtype="AIRPORT"
form="NAME"/>
<MOVELINK xml:id="mvl1"
trigger="#m1" goal="#pl1"
mover="#sne1" goal_reached="YES"/>
</isoSpace>
```

[1]We have introduced attribute-value pairs such as `type="PERSON"` for the annotion of "Mia", and also `type="FAC"` and `subtype="AIRPORT"` for that of "Jeju International Airport".

This annotation is then understood as conveying the information that there are four types of basic entities involving spatial information: *spatial named entity, motion, spatial signal*, and *place*, and that there is a relation of linking among these entities. Each entity is further specified with information provided by the assignment of a value to each relevant attribute. The place "Jeju International Airport" is, for instance, specified as FAC (facility type) being an airport. With the attribute `@target` specified as above, each of the four basic entity types `<PLACE>`, `<MOTION>`, `<SPATIAL_SIGNAL>`, and `<SPATIAL_NE>` refers to some markable (sequence of tokens) in the text.

The annotation given above then introduces one link, namely `<MOVELINK>`, among those four basic entities. This link is triggered by the motion (m1) of driving to its goal, the airport (pl1) named "Jeju International Airport", with its agent (driver) being a person (sne1) named "Mia". The link, as is annotated here, thus fully represents the information conveyed by the sentence given above. The annotation as a whole can be formally interpreted in first-order logic, as below:

(4) $\exists\{x, y, e\}[person(x) \wedge named(x, Mia) \wedge airport(y) \wedge named(y, JejuInt.Airport) \wedge move\_external(e) \wedge agent(x, e) \wedge goal(y, e) \wedge reach(x, y)]$

3 Making References to Markables

In annotating a text, each basic entity type can easily refer to a part of it as its markable because texts are considered to be sequences of character strings and can thus be tokenized. On the other hand, if input data are other than a text, then making reference to markables requires more complex processes than the simple process of segmenting a text into character offsets or tokens. In this section, we will show how making references to so-called markables in the

Figure 1: Deep Breathing ©Ghang Lee

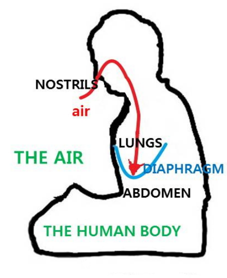

Figure 2: Deep Breathing Annotated ©Ghang Lee

annotation of non-textual data requires techniques more than simply segmenting a text.

Consider Figure 1: *Deep Breathing*. This figure is introduced as part of a guidebook for teaching how to breathe deep down to the abdomen by expanding the diaphragm during the Zen meditation. This figure cannot be segmented into character offsets or tokens, for it contains no characters at all. It rather consists of several geometric objects: (1) an area totally enclosed with a boundary line and an open area outside of it, (2) a curved line located within the enclosed area, and (3) a directed line, namely, arrow entering the upper part of the enclosed area and then reaching that curved line located at the lower part of the enclosed area.

The description of these objects may have to be more explicit for the purposes of computing, perhaps requiring the use of such notions as pixels, coordinates, orientations or topological properties to make them referable as markables. From ordinary linguistic points of view, however, such a specification seems to go beyond the level of semantic representation. It is too complicated to focus on relevant information from the given figure. Instead, we can propose a conventionally more acceptable linguistic technique. Namely, it is to assign a unique name to each of these geometric objects, thus making them uniquely identifiable within a restricted domain and producing a figure such as Figure 2: *Deep Breathing Annotated*. Such a naming technique is especially plausible because the original figure is accompanied by a title that tells what is being depicted. Because of its title *Deep Breathing*, we can conjecture that the figure depicts the process of deep breathing, sometimes called *diaphragmatic breathing*, that undergoes the expansion of the diaphragm or the abdomen.

With such knowledge, we can give names to (1) the two spatial areas: the enclosed area is named THE HUMAN BODY, that represents the shape of the human body with a sitting posture, whereas the open area outside it is named THE AIR; (2) the three relevant points: the first point is named NOSTRILS, which lies on the upper left boundary of the enclosed area, the second point, named LUNGS, which is located at the left middle part of the enclosed area, and the third point, named ABDOMEN, which is located at the mid-lower part of the same enclosed area; and (3) the two lines: the arrow is named IN_PATH, which starts from THE AIR area, goes through the NOSTRILS and the LUNGS and terminates at the ABDOMEN, whereas the other line is named DIAPHRAGM, which is shown to be stretched to the ABDOMEN. We should also be able to recognize two motions: one motion is that of an object named `air` which follows through IN_PATH, and the other motion is that of the DIAPHRAGM that expands from the LUNGS down to the ABDOMEN. Here, two moving objects, `air` and DIAPHRAGM can be treated of type SPATIAL_NE, named entities involving motions in space.

With all these names specified as above, *ISO-Space* can now be applied to the annotation of the whole figure, as represented in XLM below . Besides introducing two spatial named entities (sne1) and (sne2), it annotates two big areas, one enclosed (pl1) and the other open (pl2), the four places or points (pl3, pll4, pl5, pl6) in the enclosed area as parts of the HUMAN BODY, and a path (pl) from the open area (pl2), named THE AIR, to the ABDOMEN (pl6) involving a MOVE_IN motion of air (sne1). There are also two types of links: (1) five `QSLINK`s that relate each of the four places as well as the path to the HUMAN BODY (pl1) and (2) two MOVELINKs, one of which (mvl1) annotates the process of breathing air down to the ABDOMEN (pl6), while the other (mvl2) annotates the streching of the DIAPHRAGM (pl5) to the ABDOMEN (pl6).

(5)
```
<isoSpace xml:id="a2">
<SPATIAL_NE xml:id="sne1"
target="#figure2:air"
type="NATURAL" subtype="AIR"/>
<SPATIAL_NE xml:id="sne2"
target="#figure2:DIAPHRAGM"
type="NATURAL"/>
<PLACE xml:id="pl1"
target="#figure2:HUMAN BODY"/>
<PLACE xml:id="pl2"
target="#figure2:THE AIR area"/>
<PLACE xml:id="pl3"
target="#figure2:NOSTRILS"/>
<PLACE xml:id="pl4"
target="#figure2:LUNGS"/>
<PLACE xml:id="pl5"
target="#figure2:DIAPHRAGM"/>
<PLACE xml:id="pl6"
target="#figure2:ABDOMEN"/>
<PATH xml:id="p1"
target="#figure2:ARROW
figure" beginPoint="#pl2"
midPoint="#pl3,#pl4"
endPoint="#pl5"/>
<MOTION xml:id="m1"
motion_type="PATH"
motion_class="MOVE_INTERNALLY"/>
<MOTION xml:id="m2"
motion_type="MANNER"
motion_class="MOVE"/>
<QSLINK xml:id="qsl2"
figure="#pl2" ground="#pl1"
relType="EC(Externally
connected)"/>
<QSLINK xml:id="qsl1"
figure="#pl3" ground="#pl1"
relType="TTP(tangential proper
part)"/>
<QSLINK xml:id="qsl2"
figure="#pl4" ground="#pl1"
relType="NTTP(non-tangential
proper part)/IN"/>
<QSLINK xml:id="qsl2"
figure="#pl5" ground="#pl1"
relType="NTTP(non-tangential
proper part)/IN"/>
<QSLINK xml:id="qsl2"
figure="#pl6" ground="#pl1"
relType="NTTP(non-tangential
proper part)/IN"/>
<MOVELINK xml:id="mvl1"
trigger="#m1" source="#pl2"
goal="#pl6" mover="#sne1"
pathID="#p1" goal_reached="YES"/>
<MOVELINK xml:id="mvl2"
trigger="#m2" source="#pl5"
goal="#pl6" mover="#sne2"
goal_reached="YES"/>
</isoSpace>
```

Figure 3: Jeju Island

As is discussed in Mani and Pustejovsky (2012), the relation types such as EC, TTP, and NTTP of qualitative spatial link (QSLINK) are defined by the Region Connection Calculus 8 (RCC-8) (Randell et al., 1992) and (Galton, 2000).[2] This annotation is then understood as stating that air goes into the abdomen in the human body through the nostrils and the lungs by stretching the diaphragm, as claimed by meditation teachers.

Consider another non-textual dataset, Figure 3: *Jeju Island*.[3] This is an aerial photograph of the island. Again from the title of the figure, we understand that the oval shape refers to Jeju Island. With

[2]Here, NTTP may be replaced with IN.

[3]This is a file from the Wikimedia Commons, created by *NASA*. *http://en.wikipedia.org/wiki/File:Cheju_etm_2000097_lrg.jpg.*

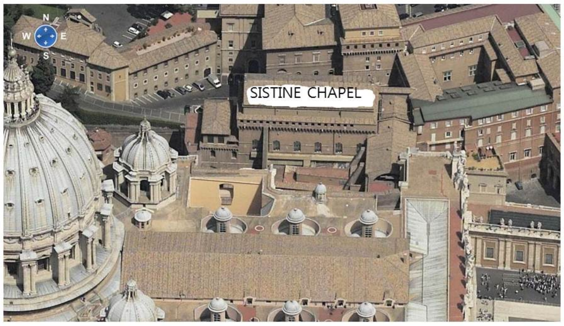

Figure 4: Sistine Chapel

Table 1: Train Schedule ©Societ Aeroporto Toscano 2002 - 2008

Aeroporto	06:53	09:03	11:03	13:03	15:03	17:03
Pisa Centrale	06:58	09:11	11:11	13:11	15:11	17:11
Pontedera	07:22					
Empoli	07:46					
Firenze SMN	08:22	10:00	12:00	14:00	16:00	18:00
NOTE/REMARKS	A	RV	RV/A	RV	RV/A	RV

A = Except on Sundays and Bank Holidays, **RV** = Fast Regional Connections

plane geometry, we may be able to define the given elliptical region or (near) convex hull and talk about its center or peripheral areas. With some knowledge of reading geographic photographs, we may also be able to derive some geographic information about its mountainous regions, surrounding oceans, attached small isles, and populated areas. We can also refer to each of those areas by drawing (Cartesian) coordinate lines, both horizontal and vertical, as detailed as necessary, over the whole photographed area, thus relying on other than linguistic knowledge or techniques such as word segmentation.[4]

In ordinary conversations, as was just claimed, we may prefer to talk about some areas with their specific names rather than their coordinate values. Naming is an important aspect of the ordinary use of language: for instance, naming places with street number, often framed in mapping coordinates, is found very useful especially when we travel to locate places. Knowing directions is also important. But photographs like Figure 3 do not have any place names or street numbers at all. It also fails to tell which is north or south and which is east or west, although they may allow us to measure a distance from one location to another. In section 5, we discuss multi-layered annotation, showing how such an approach combines various types of information, whether non-linguistic or linguistic, to enrich the annotation of non-textual data such as figures or maps. Note again that one particular layer deals with naming.

Here is a third example, Figure 4: *Sistine Chapel*. It is again an aerial photograph of St. Peter's Basilica in the Vatican with some of its surrounding buildings, one of which is the Sistine Chapel. The photograph itself would not show which building is the Sistine Chapel. The name of the chapel was later printed on the roof of its building in the photograph, Figure 4. We can thus identify the chapel as being located in the upper center of the photograph, standing just next to a smaller dome on the right of the main dome of the basilica when you enter it. Nevertheless, we still do not know how to enter it, except guessing that we might be able to enter it through the basilica. (Yes, you can, if you are a Vatican guard or dignitary.) As is again to be discussed in section 5, this photograph with the name of the destination can provide an important clue for entering the chapel only when it is annotated with other layers of information.

4 Understanding Conventions

While presenting information in a visually accessible mode, non-textual data such as maps or figures, or even textual data in a tabular form often fail to provide detailed information unless contextual information supplements them. In this section, we discuss how conventional knowledge helps interpret non-textual data.

Consider Table 1: *Train Schedule*.[5] Schedules for transportations such as trains, buses, ships, and planes are very often presented in a tabular form with columns and rows each identified. To be able to read them, however, one must know some conventions to interpret them. On the first (left-most) column five train stations are listed in order from the Aeroporto station to the Firenze SMN station, the times on each row list the departure or arrival times of trains at each station, and so on. The 09:03 train from Aeroporto stops at Pisa Central, but runs directly to Firenze without stopping at the other two

[4]If we are using a Google earth map, then we can simply rely on the geo-coordinate information provided by it.

[5]The departure times for the last two trains are deleted here.

Table 2: Flight Schedule

Ms Mia Lee			
Gimpo-Haneda	(11/30, Fri, 2012)	12:10-14:15	JL0092
Haneda-Gimpo	(12/02, Sun, 2012)	15:30-18:05	JL0093

stations in between. One gets all this information if he or she knows how to read the schedule. If one does not know about the convention of presenting such schedules for transportation, she or he may fail to get necessary information.[6]

Here is another example: a flight schedule given in a tabular form, provided by a travel agent. Knowing some conventions of printing out flight schedules, we get proper information about (1) the customer Ms Mia Lee, who was traveling from Gimpo Airport to Haneda Airport and then from Haneda back to Gimpo, (2) the respective departure and arrival dates and times of the on-going and return flights, and (3) the names of the carriers.

With such knowledge, we can annotate this table with *ISO-Space*, as shown below.

(6)
```
<isoSpace xml:id="a3">
<SPATIAL_NE xml:id="sne1"
target="#table2:col1,row1:
[token1,token3]" form="NAME"
type="PERSON"/>
<SPATIAL_NE xml:id="sne2"
target="#table2:col4,row2:token1"
form="NAME" type="PLANE"
subtype/flightNo="JL0092"/>
<PLACE xml:id="pl1"
target="#table2:col1,row2:token1"
from="NAME" type="FAC"
subtype="AIRPORT" city="SEOUL"
country="KR"/>
<PLACE xml:id="pl2"
target="#table2:col1,row2:token3"
from="NAME" type="FAC"
subtype="AIRPORT" city="TOKYO"
country="JP"/>
<PATH xml:id="p1"
begingPoint="#pl1"
endPoint="#pl2"/>
<MOTION xml:id="m1"
motion_type="MANNER"
motion_class="LEAVE"/>
<MOTION xml:id="m2"
motion_type="MANNER"
motion_class="REACH"/>
<MOTION xml:id="m3"
motion_type="MANNER"
motion_class="MOVE_EXTERNAL"/>
<MOVELINK xml:id="mvl1"
trigger="#m1,#m2" mover="#sne1"
means="#sne2" source="#pl1"
goal="#pl2" goal_reached="YES"
pathID="#p1"/>
</isoSpace>
```

[6]Strcitly speaking, these tables are only partially *non-textual*. They are non-textual in the sense that they are laid out differently from the ordinary text data.

Here three <MOTION> elements are not anchored at all, but only understood through some conventional knowledge involving air flights. These elements should be introduced in order to be able to annotate the departure and arrival-related spatio-temporal information provided in the second and third rows of table 2.

As can be noted very easily, *ISO-Space* deals with spatial information only. To annotate temporal information, it should be applied jointly with *ISO-TimeML* (2012). We can then make the example more interesting and sensible, by annotating various quantitative information of spatio-temporal measurements such as time amount, durations, frequency, distance, and also the tense and modal property of motions or events in general. Lee (2012) has already argued that such a joint application is possible because both *ISO-Space* and *ISO-TimeML* are designed to be interoperable.

5 Multi-layered Annotation

As is argued by Berg et al. (2010) and is well proven by Google Earth map resources, no single map can provide all of the necessary geographic information, thus requiring several layers of a map. If a single map is marked up with all the information, it cannot be parsed. On the other hand, if it is just an aerial photograph, it may not contain enough information, for instance, to tell which town is which and which road is which. This could be the case with linguistic annotation, too. If a single text is tokenized and annotated with all sorts of grammatical or seman-

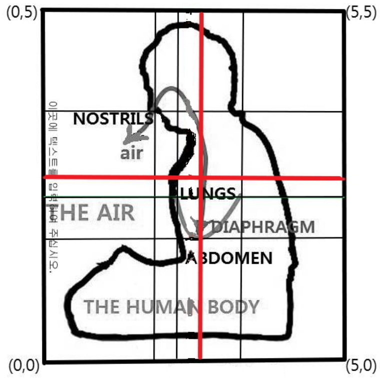

Figure 5: Deep Breathing Figure Segmented ©Ghang Lee

Figure 6: Jeju Island-annotated

tic information, all the information may be too tangled up to be retrieved. LAF (2012) thus requires standoff annotation, as opposed to in-line annotation, while allowing multi-layered annotation of linguistic information. Accordingly, we also argue that a multi-layered approach is not only suitable, but required for the annotation of non-textual data as well as textual data.

For illustration, consider again the figures of deep breathing. In section 3, we have discussed two figures, Figure 1:*Deep Breathing* and Figure 2:*Deep Breathing Annotated*. We have then argued how *ISO-Space* can be adequately applied to annotate the figure of deep breathing by making references to the entity names specified in the second figure. Nevertheless, one may argue that naming alone is not fine-grained enough to identify regions and other spatial entities for some technical applications such as drawing cartoons or architectural designs or even annotating them. In addition to the technique of naming, we thus propose another technique as providing an additional layer of making it possible to refer to markables in both textual and non-textual data.

This technique is a well-known technique of segmenting data, whether textual or not, into smaller constituents. Just like maps with geo-coordinates, each (two-dimensional) figure in a text is to be treated like a Cartesian plane, divided into small areas with their coordinates specified.[7] Then the character strings and some defining points of the region or its parts such as the nostrils, the lungs, the diaphragm, and the abdomen should be identified strictly in terms of those coordinates, just as a text is segmented into tokens based on character offsets.

This technique can be illustrated with the figure of deep breathing. In addition to those two figures, introduced in 3, we can introduce one more figure, Figure 5: *Deep Breathing Figure Segmented*. This third figure treats the whole region as a two-dimensional Cartesian plane, segmented into 5 x 5 areas with unequal sizes.[8] Horizontal and vertical lines are drawn in such a way that some relevant points can be identified with some of their intersections. The position of the nostrils, for instance, is identified with the point (1,4). The non-stretched diaphragm can also be identified as a line segment from (2,2) to (4,2), while its mid-point is being stretched to the point (3,1). Likewise, all of the relevant areas can also be identified by drawing additional lines, if necessary, that segment the whole area into much smaller areas. This then requires another layer of representing the whole figure.

For another illustration, consider the following map of Jeju Island, Figure 6: *Jeju Island-annotated*.[9] Unlike the aerial photograph of Jeju, Figure 3, this new figure has names for several locations: (1) Mt. Halla for the mountain located in the center of the island, (2) Jeju City, Seogwipo, and Jungmun Resort for three populated areas, and Jeju International

[7] Geo-coordinates or other map reading coordinates are particular instances of the Cartesian coordinate.

[8] Quantative information is irrelevant for this particular example.

[9] This file is copyrighted by Jeju Special Self-Governing Province. ©Jejumaster@juju.go.kr. The red line, indicating the Pyeonghwa Route, is added by the author.

Figure 7: Jeju Google Earth

Airport for the airport, and Pyeonghwa Route for a highway mostly connecting the airport and Jungmun Resort. The two figures offer different types of geographic information: Figure 3 shows the elevation of each part of the island, while Figure 6 provides information more for traveling around the cities on the island. In Figure 6, there is a little arrow on the leftmost upper corner pointing to the north, providing directional information. With this information, we know that the airport is located in the north central boundary of the island. Combined together, these two figures can provide a lot of information that we may or may not be able to derive from a text alone.

Here is a map of the same island, Figure 7 *Jeju Google Earth*[10], with the old romanized name "Cheju do" of the Jeju Province.[11] Besides some place names printed on it, the map contains a lot of tiny buttons, either square-shaped or camera-shaped. As any of the buttons is kept being clicked, it keeps displaying different layers of the map with more detailed information, texts or photos. The Google earth map is thus a typical example of displaying information in layers.

Finally, consider a map for the Sistine Chapel in the Vatican, Figure 8: *How to Get to the Sistine Chapel*.[12] This map guides one from Piazza Pio XII,

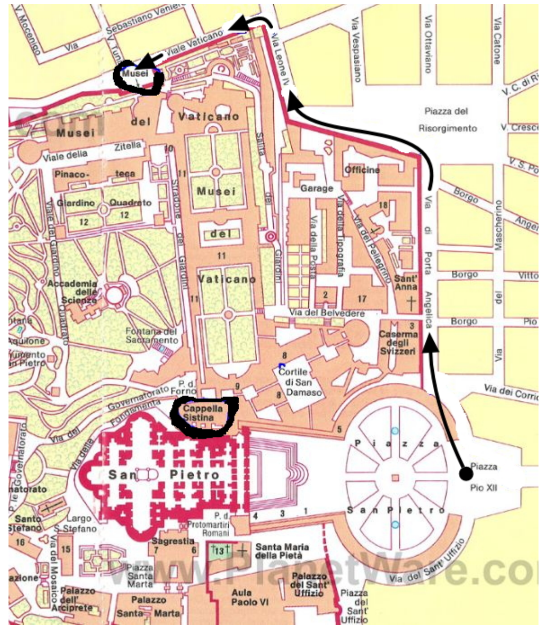

Figure 8: How to Get to the Sistine Chapel

which is just at the entrance to Piazza San Pietro, to the Vatican Museums following the sequence of the arrows going through the roads, named *Via di Porta Angelica*, *Via Leone IV*, and *Viate Vaticano*. There she could enter the museums and all the way to the chapel, named *Capella Sistina* in Italian.

6 Concluding Remarks

This paper applies *ISO-Space* to the annotation of non-textual data such as maps and figures or even some textual data presented in a tabular form because spatial information is very often carried by such data. In annotating such data, one difficulty was how to anchor such basic entities as `PLACE` and `PATH` to parts of the data, since pictures and figures, unlike texts, cannot be tokenized. Another difficulty arose from the understanding of various symbols or conventional cues in visual data. A non-location entity `MOTION` of *ISO-Space*, for instance, is seldom mentioned explicitly, but only expressed implicitly

[10]Created by U.S. Department of State Geographer, ©2013 Google, ©2009 Geo-Basis-DE/BKG, DATA SIO, NOAA, U.S. Navy, NGA, GEBCO.

[11]The place name "Jeju-do" is ambiguous: it may mean either the *island* or the *province* of Jeju. The name "Jeju" itself is also ambiguous: it may refer to either the *city* or the *province* of Jeju.

[12]Except for the paths to the entrance to the Vatican Museums marked by the author, this map is provided by PlanetWare.com with the following note: Use this map on your web site - copy and paste the code below: `<a href="http://www.planetware.com/map/vatican-city-map-scv-vat_ce.htm"> <img src ="http://www.planetware.com/i/map/SCV/vatican-city-map.jpg" width="1200" height="899"/></a><br/> Map from <a href="http://www.planetware.com/">PlanetWare.com</a>.`

with a little pointed arrow, as in Figure 1 or Figure 8. We have argued that such difficulties can be overcome if different layers of visual data are presented and also if various types of information from those data are combined in a consistent way. We have also proposed two conventional techniques for the treatment of markables in annotation: one is to name relevant elements in non-textual data and another is to segment figures in a referable way, for instance, with coordinates. Naming and segmentation are then shown to be providing different layers of annotation, as needs arise.

We have, however, treated these issues simply as technical issues for linguistic purposes only. We have thus avoided discussing any theoretical implications that may go beyond the domain of linguistic annotation, although we have not explicitly demarcated the line between what is linguistic and what is not. A question still remains whether the annotation of non-textual data or multimedia is part of linguistic work. For computing purposes, however, more serious questions may be raised. One could ask how non-human agents can annotate such non-textual data for spatial or spatio-temporal information. Towards answering these questions, more work should be done on multimedia or motion tagging, as discussed in Mani and Pustejovsky (2012), and more serious references should be made to some initiatives that exist in GIS(Geographic Information System)-related communities.

Acknowledgements

I own many thanks to Suk-jin Chang, Jae-Woong Choe, Roland Hausser, Hwan-Mook Lee, Ghang Lee, Chongwon Park, and four anonymous reviewers for their very constructive and detailed comments that helped improve the paper.

References

Berg, Mark de, Otfried Cheong, Marc van Kreveld and Mark Overmars. 2010. *Computational Geometry: Algorithms and Applications*, 3rd edition. Springer, Berlin.

Bunt, Harry. 2010. A methodology for designing semantic annotation languages exploiting syntactic-semantic iso-morphisms. In: A. Fang, N. Ide and J. Webster (eds.) *Proceedings of ICGL 2010, the Second International Conference on Global Interoperability for Language Resources*, pp. 29-45. Hong Kong City University.

Bunt, Harry. 2011. Introducing abstract syntax + semantics in semantic annotation, and its consequences for the annotation of time and events. In E. Lee and A. Yoon (eds.), *Recent Trends in Language and Knowledge Processing*, pp. 157-204. Hankukmunhwasa, Seoul.

Galton, Antony. 2000. *Qualitative Spatial Change*. Oxford University Press, Oxford.

ISO 24612:2012(E) *Language resource management - Linguistic annotation framework (LAF)*, International Organization for Standardizations, Geneva.

ISO 24617-1:2012(E) *Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML)*. International Organization for Standardizations, Geneva.

Lee, Kiyong. 2012. Towards interoperable spatial and temporal annotation schemes. *Proceedings of the Joint ISA-7, SRSL-3, and I2MRT Workshop on Semantic Annotation and the Integration and Interoperability of Multimodal Resources and Tools*, a satellite workshop (26-27 May 2012) held in conjunction with LREC 2012. Istanbul.

Mani, Inderjeet, and James Pustejovsky. 2012. *Interpreting Motion: Grounded Representations for Spatial Language*. Oxford University Press, Oxford.

Pustejovsky, James, Jessica Moszkowics, and Marc Verhagen. 2012. The current status of ISO-Space. *Proceedings of the Seventh Workshop on Interoperable Semantic Annotation (ISA-7)*, a satellite workshop held in conjunction with LREC 2012. Istanbul.

Randell, David A., Zahn Cui, and Anthony G. Cohn. 1992. A spatial logic based on regions and connection. *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, pp. 165-175. Morgan Kaufman, San Mateo, CA.