The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

Intuitive spatiotemporal representation based on Mental Image Description Language L_{md}

Masao Yokota

Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka-shi, 811-0295, Japan
(Tel : 81-92-606-5897; Fax : 81-92-606-8923)
(yokota@fit.ac.jp)

Abstract: Mental Image Directed Semantic Theory (MIDST) has defined the semantic content (i.e., concept) of a spatiotemporal expression as a certain generalized mental image of its referents in the physical world and proposed a method to model mental images as loci in attribute spaces so called. These loci correspond with events in certain attributes of physical matters and are described in a formal language L_{md}, so called, "Mental Image Description Language" employed for many-sorted predicate logic. This paper presents a brief sketch of L_{md} and a systematic method to formulate and compute natural concepts of physical reality comprising spatiotemporal language semantics in order to facilitate intuitive human-system interaction.

Keywords: Spatiotemporal knowledge, Knowledge representation language, Mental image model.

I. INTRODUCTION

Most of the traditional approaches to spatial language understanding have focused on computing purely objective geometric relations (i.e., topological, directional and metric relations) conceptualized as spatial prepositions or so, considering properties and functions of the objects involved [1-3]. From the semantic viewpoint, spatial expressions have the virtue of relating in some way to visual scenes being described. Therefore, their semantic descriptions can be grounded in perceptual representations, possibly, cognitively inspired and coping with all kinds of spatial expressions including such verb-centered ones as S1 and S2 as well as preposition-centered ones.

(S1) The path *sinks* to the brook.

(S2) The path *rises* from the brook.

These verb-centered expressions are assumed to reflect not much the purely objective geometrical relations but very much certain dynamism at human perception of the objects involved because they can refer to the same scene in the external. This is also the case for S3 and S4.

(S3) The roads *meet* there.

(S4) The roads *separate* there.

The Mental Image Directed Semantic Theory (MIDST) [4] has proposed a dynamic model of human perception yielding omnisensory image of the world and classified natural event concepts (i.e., event concepts in natural language) into two types of categories, 'Temporal Events' and 'Spatial Events'. These are defined as temporal and spatial changes (or constancies) in certain attributes of physical objects, respectively, with S1 and S2 included in the latter. Both the types of events are uniformly analyzable as temporally parameterized loci in attribute spaces and describable in a formal language L_{md}.

This language is employed for many-sorted first-order predicate logic and can provide spatiotemporal expressions with computable semantic descriptions as their perceptual representations. The most remarkable feature of L_{md} is its capability of formalizing spatiotemporal event concepts on the level of human sensations while the other similar knowledge representation languages are designed to describe the logical relations among conceptual primitives represented by lexical tokens [5,6].

This paper presents a brief sketch of L_{md} and a systematic method to formulate and compute natural concepts of physical reality comprising spatial language semantics in order to facilitate integrated spatial language understanding.

II. FORMAL LANGUAGE L_{md}

In the MIDST, word meanings are treated in association with mental images, not limited to visual but omnisensory, modeled as "Loci in Attribute Spaces". An attribute space corresponds with a certain measuring instrument just like a barometer, thermometer or so and the loci represent the movements of its indicator.

For example, the moving gray triangular object shown in Fig.1-Left is assumed to be perceived as the loci in the three attribute spaces, namely, those of 'Location', 'Color' and 'Shape' in the observer's brain. A general locus is to be articulated by "Atomic Locus" with the duration $[t_i, t_f]$ as depicted in Fig.1-Right and formulated as (1).

$$L(x,y,p,q,a,g,k) \qquad (1)$$

This is a formula in many-sorted predicate logic, where "L" is a predicate constant with five types of terms: "Matter" (at 'x' and 'y'), "Attribute Value" (at 'p' and 'q'), "Attribute" (at 'a'), "Event Type" (at 'g') and "Standard" (at 'k'). Conventionally, Matter variables are headed by 'x', 'y' and 'z'. This formula is called

'Atomic Locus Formula' whose first two arguments are sometimes referred to as 'Event Causer (EC)' and 'Attribute Carrier (AC)', respectively while ECs are often optional in natural concepts such as intransitive verbs. For simplicity, the syntax of L_{md} allows Matter terms (e.g., 'Tokyo' and 'Osaka' in (2) and (3)) to appear at Attribute Values or Standard in order to represent their values at the time. Moreover, when it is not so significant to discern ECs or Standards, anonymous variables, usually symbolized as '_', can be employed in their places (See (22) for example).

The intuitive interpretation of (1) is given as follows. *"Matter 'x' causes Attribute 'a' of Matter 'y' to keep (p=q) or change (p ≠ q) its values temporally (g=Gt) or spatially (g=Gs) over a time-interval, where the values 'p' and 'q' are relative to the standard 'k'."*

When *g=Gt* and *g=Gs*, the locus indicates monotonic change or constancy of the attribute in time domain and that in space domain, respectively. The former is called 'temporal event' and the latter, 'spatial event'. For example, the motion of the 'bus' represented by S5 is a temporal event and the ranging or extension of the 'road' by S6 is a spatial event whose meanings or concepts are formulated as (2) and (3), respectively, where 'A12' denotes the attribute 'Physical Location'. These two formulas are different only at the term 'Event Type'.

(S5) The bus runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A12,Gt,k) \wedge bus(y) \quad (2)$$

(S6) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A12,Gs,k) \wedge road(y) \quad (3)$$

It has been often argued that human active sensing processes may affect perception and in turn conceptualization and recognition of the physical world.

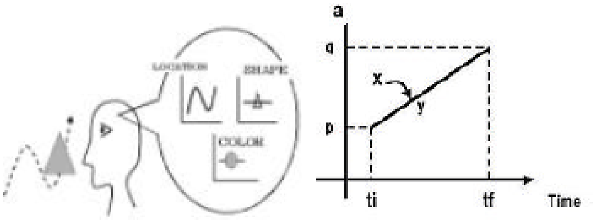

Fig.1. Mental image model (Left) and Atomic Locus in Attribute Space (Right).

The difference between temporal and spatial event concepts can be attributed to the relationship between the Attribute Carrier (AC) and the Focus of the Attention of the Observer (FAO). To be brief, the FAO is fixed on the whole AC in a temporal event but *runs* about on the AC in a spatial event. Consequently, as shown in Fig.2, the *bus* and the FAO move together in the case of S5 while the FAO solely moves along the *road* in the case of S6. That is, *all loci in attribute spaces correspond one to one with movements or, more generally, temporal events of the FAO.*

The duration of an atomic locus, suppressed in the atomic locus formula, corresponds to the time-interval over which the FAO is put on the corresponding phenomenon outside. The MIDST has employed 'tempo-logical connectives (TLCs)' representing both logical and temporal relations between loci. Articulated loci are combined with tempo-logical conjunctions, where 'SAND ($\wedge_0$)' and 'CAND ($\wedge_1$)' are most frequently utilized, standing for 'Simultaneous AND' and 'Consecutive AND', conventionally symbolized as 'Π' and '•', respectively.

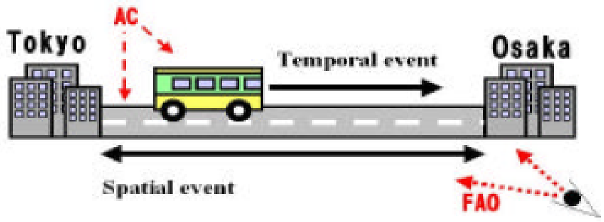

Fig.2. FAO movements and Event types.

III. EVENT CONCEPT DESCRIPTION

For example, the English verbs 'carry' and 'return' refer certain temporal events and their concepts can be defined as (4) and (5), respectively. These formulas can be depicted as Fig.3-Left and Right.

$$(\lambda x,y)carry(x,y) \leftrightarrow (\lambda x,y)(\exists p,q,k)L(x,x,p,q,A12,Gt,k)\Pi L(x,y,p,q,A12,Gt,k) \wedge x \neq y \wedge p \neq q \quad (4)$$

$$(\lambda x)return(x) \leftrightarrow (\lambda x)(\exists p,q,k)L(x,x,p,q,A12,Gt,k) \bullet L(x,x,p,q,A12,Gt,k) \wedge x \neq y \wedge p \neq q \quad (5)$$

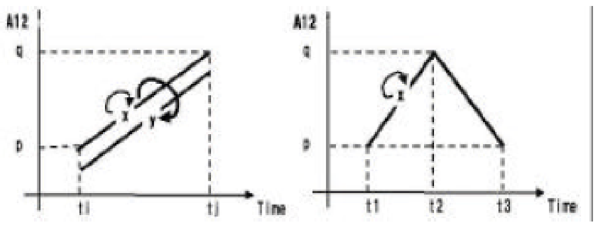

Fig.3. Loci of 'carry' (Left) and 'return' (Right).

The expression (6) is the definition of the English verb concept 'fetch' depicted as Fig.4-Right. This implies such a temporal event that 'x' goes for 'y' and then comes back with it.

$$(\lambda x,y)fetch(x,y) \leftrightarrow (\lambda x,y)(\exists p1,p2,k)L(x,x,p1,p2,A12,Gt,k) \bullet ((\underline{L(x,x,p2,p1,A12,Gt,k)\Pi L(x,y,p2,p1,A12,Gt,k)}) \wedge x \neq y \wedge p1 \neq p2 \quad (6)$$

In the same way, the English verb concept 'hand' or 'receive' depicted as Fig.4-Right is defined uniformly as (7) or its abbreviation (7').

$$(\lambda x,y,z)hand(x,y,z) \leftrightarrow (\lambda x,y,z)receive(z,,y,x) \leftrightarrow (\lambda x,y,z)(\exists k)L(x,y,x,z,A12,Gt,k)\Pi L(z,y,x,z,A12,Gt,k) \wedge x \neq y \wedge y \neq z \wedge z \neq x \quad (7)$$

$$(\lambda x,y,z)(\exists k)L(\{x,z\},y,x,z,A12,Gt,k) \wedge x \neq y \wedge y \neq z \wedge z \neq x \quad (7')$$

Such locus formulas as correspond with natural event concepts are called 'Event Patterns' and about 40 kinds of event patterns have been found concerning the attribute 'Physical Location (A12)', for example, *start, stop, meet, separate, carry, return*, etc [7].

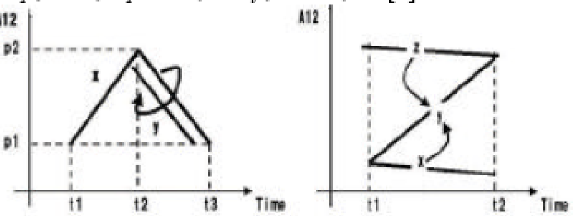

Fig.4. Loci of 'fetch' (Left) and 'hand/receive' (Right).

The attribute spaces for humans correspond to the sensory receptive fields in their brains. At present, about 50 attributes concerning the physical world have been extracted exclusively from English and Japanese words as shown in Table 1. They are associated with all of the 5 senses (i.e. sight, hearing, smell, taste and feeling) in our everyday life while those for information media other than languages correspond to limited senses. For example, those for pictorial media, marked with '*' in Table 1, associate limitedly with the sense 'sight' as a matter of course. This kind of classification of attributes plays a very important role in our cross-media operating system [7]. Correspondingly, six categories of standards shown in Table 2 have been extracted that are assumed necessary for representing values of each attribute in Table 1. In general, the attribute values represented by words are relative to certain standards as explained briefly in Table 2. For example, (8) and (9) are different formulations of a locus due to the different standards 'K_1' and 'K_2' for scaling as shown in Fig.5-Left and Right, respectively. That is, whether the point (t2, q) is significant or not, more generally, how to articulate a locus depends on the precisions or the granularities of these standards, which can be formulated as (10) and (11), so called, ***'Postulate of Arbitrariness in Locus Articulation'***. This postulate affects the process of conceptualization on a word based on its referents in the world [4].

$$L(y,x,p,q,a,g,K_1)\bullet L(y,x,q,r,a,g,K_1) \quad (8)$$

$$L(y,x,p,r,a,g,K_2) \quad (9)$$

$$(\forall p,q,r,k)(L(y,x,p,q,a,g,k)\bullet L(y,x,q,r,a,g,k).\supset.(\exists k')L(y,x,p,r,a,g,k')\wedge k'\neq k) \quad (10)$$

$$(\forall p,r,k)(L(y,x,p,r,a,g,k).\supset.(\exists q,k')L(y,x,p,q,a,g,k')\bullet L(y,x,q,r,a,g,k')\wedge k'\neq k) \quad (11)$$

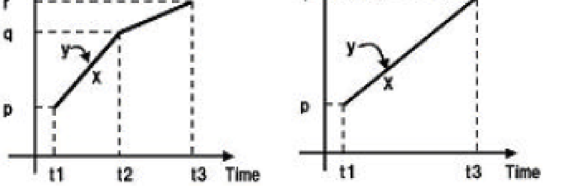

Fig.5. Arbitrariness in locus articulation due to standards: Standard k_1 (Left) is finer than k_2 (Right).

Table 1. Examples of attributes

Codes	Attribute [Property]†
*A01	PLACE OF EXISTENCE [N]
	....
*A11	SHAPE [N]
*A12	PHYSICAL LOCATION [N]
*A13	DIRECTION [N]
*A14	ORIENTATION [N]
*A15	TRAJECTORY [N]
*A16	VELOCITY [S]
*A17	MILEAGE [S]
A18	STRENGTH OF EFFECT [S]
A19	DIRECTION OF EFFECT [N]
	....
A28	TEMPERATURE [S]
A29	TASTE [N]
A30	ODOUR [N]
A31	SOUND [N]
*A32	COLOR [N]
A33	INTERNAL SENSATION [N]
A34	TIME POINT [S]
A35	DURATION [S]
A36	NUMBER [S]
A37	ORDER [S]
A38	FREQUENCY [S]
A39	VITALITY [S]
	....
*A44	TOPOLOGY [N]
*A45	ANGULARITY [S]
	....

† [S] and [N] mean "Scalor" and "Non-scalor", respectively.

Table 2. List of standards.

Categories	Remarks
Rigid Standard	Objective standards such as denoted by measuring *units* (meter, gram, etc.).
Species Standard	The *values ordinary* for a species. A *short train* is ordinarily longer than a *long pencil*.
Proportional Standard	'*Oblong*' means that the width is greater than the height at a physical object.
Individual Standard	*Much* money for one person can be too *little* for another.
Purposive Standard	One room large enough for a person's *sleeping* must be too small for his *jogging*.
Declarative Standard	The origin of an order such as 'next' must be declared explicitly just as 'next *to him*'.

As already mentioned, ***all loci in attribute spaces correspond one to one with movements or, more generally, temporal events of the FAO. This implies that L_{md} expression can suggest a robot what and how should be attended to in its environment.*** And this is why S1 and S2 can refer to the same scene in spite of their appearances, where what 'sinks' or 'rises' is the FAO and whose conceptual descriptions are given as (12) and (13), respectively, where 'A13', '↑' and '↓' refer to the attribute 'Direction' and its values 'upward' and 'downward', respectively. Such a fact is generalized as ' ***Postulate of Reversibility of a Spatial Event***

(PRS)' that can be one of the principal inference rules belonging to people's common-sense knowledge about geography. This postulation is also valid for such a pair of S7 and S8 as interpreted approximately into (14) and (15), respectively. These pairs of conceptual descriptions are called ***equivalent in the PRS***, and the paired sentences are treated as ***paraphrases*** each other.

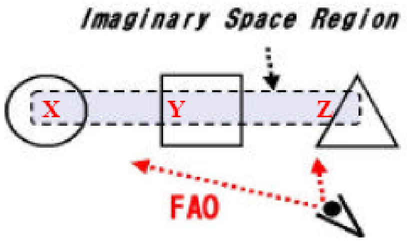

Fig.6. Spatial event 'row' and FAO movement.

$(\exists y,p,z)L(\_,y,p,z,A12,Gs,\_)\Pi L(\_,y,\downarrow,\downarrow,A13,Gs,\_)$
$\wedge path(y) \wedge brook(z) \wedge p \neq z$ (12)

$(\exists y,p,z)L(\_,y,z,p,A12,Gs,\_)\Pi L(\_,y,\uparrow,\uparrow,A13,Gs,\_)$
$\wedge path(y) \wedge brook(z) \wedge p \neq z$ (13)

(S7) Route A and Route B meet at the city.

$(\exists p,y,q)L(\_,Route\_A,p,y,A12,Gs,\_)\Pi$
$L(\_,Route\_B,q,y,A12,Gs,\_) \wedge city(y) \wedge p \neq q$ (14)

(S8) Route A and Route B separate at the city.

$(\exists p,y,q)L(\_,Route\_A,y,p,A12,Gs,\_)\Pi$
$(\_,Route\_B,y,q,A12,Gs,\_) \wedge city(y) \wedge p \neq q$ (15)

For another example of spatial event, Fig.6 concerns the perception of the formation of multiple objects, where FAO runs along an imaginary object so called 'Imaginary Space Region (ISR)'. This spatial event can be verbalized as S9 using the preposition 'between' and formulated as (16) or (16'), corresponding also to such concepts as 'row', 'line-up', etc. Employing ISRs and the 9-intersection model [3], all the topological relations between two objects can be formulated in such expressions as (17) or (17') for S10, and (18) for S11, where '*In*', '*Cont*' and '*Dis*' are the values 'inside', 'contains' and 'disjoint' of the attribute 'Topology (A44)' with the standard '9-intersection model (*9IM*)', respectively.

(S9) Y is between X and Z.

$(\exists y,p,q)(L(\_,y,X,Y,A12,Gs,\_)\Pi$
$L(\_,y,p,p,A13,Gs,\_))\bullet(L(\_,y,Y,Z,A12,Gs,\_)$
$\Pi L(\_,y,q,q,A13,Gs,\_)) \wedge ISR(y) \wedge p=q$ (16)

$(\exists y,p)(L(\_,y,Z,Y,A12,Gs,\_)\bullet L(\_,y,Y,X,A12,Gs,\_))$
$\Pi L(\_,y,p,p,A13,Gs,\_) \wedge ISR(y)$ (16')

(S10) Tom is in the room.

$(\exists x,y)L(Tom,x,y,Tom,A12,Gs,\_)\Pi$
$L(Tom,x,In,In,A44,Gt,9IM) \wedge ISR(x) \wedge room(y)$ (17)

$(\exists x,y)L(Tom,x,Tom,y,A12,Gs,\_)\Pi L(Tom,x,Cont,Cont,$
$A44,Gt,9IM) \wedge ISR(x) \wedge room(y)$ (17')

(S11) Tom exits the room.

$(\exists x,y,p,q)L(Tom,Tom,p,q,A12,Gt,\_)\Pi L(Tom,x,y,Tom,A$
$12,Gs,\_)\Pi L(Tom,x,In,Dis,A44,Gt,9IM)$
$\wedge ISR(x) \wedge room(y) \wedge p \neq q$ (18)

IV. CONCLUSION

Table 1 and 2 show that ordinary people live their casual lives, attending to tens of attributes of the matters in the world to cognize them in comparison with several kinds of standards. The conceptual description of an event in L_{md} is compared to a movie film recorded through a floating camera because it is necessarily grounded in FAO's movement over the event. The author has analyzed a considerable number of spatial event terms over various kinds of English words such as prepositions, verbs, adverbs, etc. categorized as 'Dimensions', 'Form' and 'Motion' in the class 'SPACE' of the Roget's thesaurus [7], and found that the greater part of the concepts of spatial event terms can be defined in exclusive use of five kinds of attributes for FAOs, namely, 'Physical location (A12)', 'Direction (A13)', 'Trajectory (A15)', 'Mileage (A17)' and 'Topology (A44)'. This fact implies that L_{md} expression can control robotic attention mechanism efficiently in a top-down way. At my best knowledge, there is no other theory or method [e.g., 6,9] that can provide spatiotemporal expressions with semantic interpretation in such a systematic way where both temporal and spatial events are simply and adequately formulated by controlling the term of Event Type of the atomic locus formula reflecting FAO movement.

REFERENCES

[1] Logan G.D & Sadler DD (1996), A computational analysis of the apprehension of spatial relations. In Bloom P, Peterson MA, Nadel L & Garrett M (Eds.), Language and Space, Cambridge, MA: MIT Press 493-529

[2] Coventry KR., Prat-Sala M & Richards LV (2001) The interplay between geometry and function in the comprehension of 'over', 'under', 'above' and 'below'. Journal of Memory and Language, 4:376-398

[3] Shariff AR, Egenhofer M & Mark D (1998), Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English-language Terms. International Journal of Geographical Information Science, 12-3:215-246.

[4] Yokota M (2005), An Approach to Integrated Spatial Language Understanding Based on Mental Image Directed Semantic Theory. Proc. of 5th Workshop on Language and Space, Bremen, Germany, Oct.

[5] Sowa JF (2000), Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA

[6] Miller GA & Johnson-Laird PN (1976), Language and Perception, Harvard University Press

[7] Roget P (1975), Thesaurus of English Words and Phrases, J.M.Dent & Sons Ltd., London

[8] Yokota M & Capi G (2005), Cross-media Operations between Text and Picture Based on Mental Image Directed Semantic Theory. WSEAS Trans. on Information Science and Applications, 10-2:1541-1550

[9] Langacker R (1991), Concept, Image and Symbol. Mouton de Gruyter, Berlin/New York