An Approach to Integrated Spatial Language Understanding Based on Mental Image Directed Semantic Theory

Masao Yokota

Department of System Management, Faculty of Information Engineering, Fukuoka Institute of Technology

Abstract

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. The Mental Image Directed Semantic Theory (Yokota, et al, 1991) has proposed an omnisensory mental image model and its description language $\boldsymbol{L_{md}}$. This language is employed for many-sorted predicate logic and can provide spatial expressions with computable semantic descriptions as their perceptual representations. This paper presents a brief sketch of the MIDST, and focuses on word meaning description and spatial language understanding in association with the mental image model for human-robot dialogue facilitation.

1. Introduction

Most approaches to spatial language understanding have focused on computing geometric relations (i.e., topological, directional and metric relations) conceptualized as spatial prepositions, considering properties and functions of the objects involved (e.g., Logan & Sadler, 1996; Coventry, Prat-Sala & Richards, 2001).

For more precise and integrated spatial language understanding, however, individual words or collocations concerned should be provided with more precise semantic definitions in some systematically computable form. The Mental Image Directed Semantic Theory (MIDST) has proposed an omnisensory mental image model and its intermediate description language $\boldsymbol{L_{md}}$ (Yokota et al, 1991; Yokota, 2005). This language is employed for

many-sorted first-order predicate logic and can provide spatio-temporal expressions with computable semantic descriptions as their perceptual representations. In the MIDST, the concepts conveyed by such syntactic components as words, phrases, clauses and so on are associated with mental imagery of the external or physical world and intermediately formalized in $\boldsymbol{L_{md}}$. The most remarkable feature of $\boldsymbol{L_{md}}$ is its capability of formalizing spatio-temporal 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 (e.g., Sowa, 2000; Miller & Johnson-Laird, 1976).

2. Brief sketch of MIDST

2.1 Omnisensory image model

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, a map measurer or so and the loci represent the movements of its indicator.

For example, the moving grey triangular object shown in Fig.1 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.2-a and formulated as (1).

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

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.

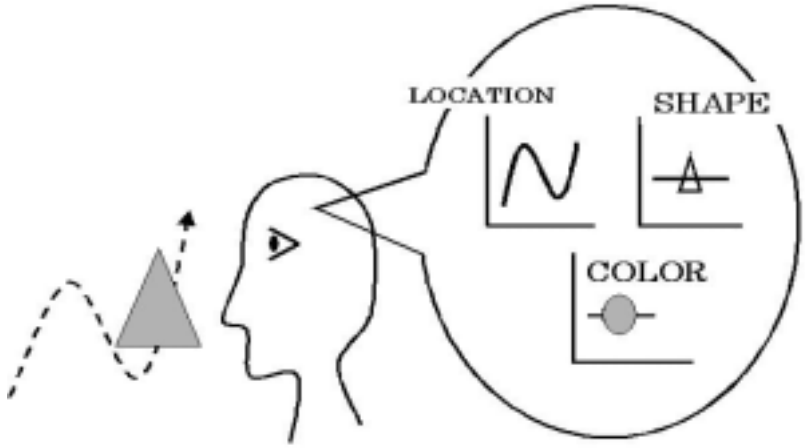
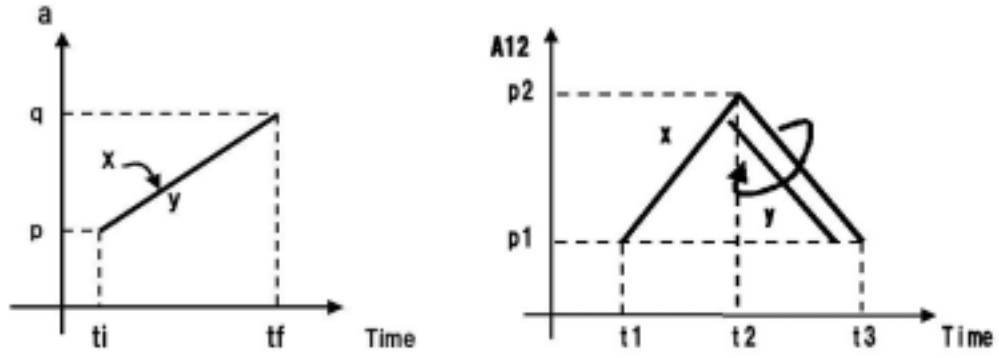


Fig.1. Mental image model.



(a) **(b)**

Fig.2. Atomic locus (a) and Locus of 'fetch' (b).

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 S1 is a temporal event and the ranging or extension of the 'road' by S2 is a spatial event whose meanings or concepts are formulated as (2) and (3), respectively, where

the attribute is 'Physical Location' denoted by 'A12'. For simplicity, Matter terms (e.g., 'Tokyo' and 'Osaka' in S1 and S2) are often placed at Attribute Values (See (45) in Section 5) or Standard to represent their values at the time.

(S1) The bus runs from Tokyo to Osaka.

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

(S2) The road runs from Tokyo to Osaka.

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

2.2 Tempo-logical connectives

The duration of an atomic locus, suppressed in the atomic locus formula, corresponds to the time-interval over which the Focus of the Attention of the Observer (FAO) is put on the corresponding phenomenon outside. The MIDST has employed 'tempo-logical connectives' representing both logical and temporal relations between loci.

A tempo-logical connective K_i is defined by (4), where τ_i, χ and K refer to one of the temporal relations indexed by *'i'*, a locus, and an ordinary binary logical connective such as the conjunctive '$\wedge$', respectively. This is more natural and economical than explicit indication of time intervals, considering that people do not consult chronometers all the time in their daily lives.

$$\chi_1 K_i \chi_2 \leftrightarrow (\chi_1 K \chi_2) \wedge \tau_i(\chi_1, \chi_2) \quad (4)$$

The expression (5) is the conceptual description of the English verb 'fetch' depicted as Fig.2-b, implying such a temporal event that 'x' goes for 'y' and then comes back with it, where 'Π'and '•' are tempo-logical connectives, 'SAND' and 'CAND', standing for 'Simultaneous AND' and 'Consecutive AND', respectively.

$$(\exists x,y,p1,p2,k)L(x,x,p1,p2,A12,Gt,k)\bullet((L(x,x,p2,p1,A12,Gt,k)\Pi L(x,y,p2,p1,A12,Gt,k))\wedge x\neq y\wedge p1\neq p2 \quad (5)$$

Such an expression as (5) is called ‘Event Pattern’ and about 40 kinds of event patterns have been found concerning the attribute ‘Physical Location (A12)’, for example, *start*, *stop*, *meet*, *separate*, *return*, etc (Yokota, et al, 1991).

Furthermore, a very important concept called ‘Empty Event (EE)’ and denoted by ‘ε’ is introduced. An EE stands for nothing but for time elapsing and is explicitly defined as (6) with the attribute ‘Time Point (A34)’. According to this scheme, the duration [*t1, t2*] of an arbitrary locus X can be expressed as (7).

$$\varepsilon \leftrightarrow (\exists x,y,t1,t2,g,k)\ L(x,y,t1,t2,A34,g,k) \quad (6)$$

$$X\ \Pi\ \varepsilon(t1, t2) \quad (7)$$

2.3 Event types

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

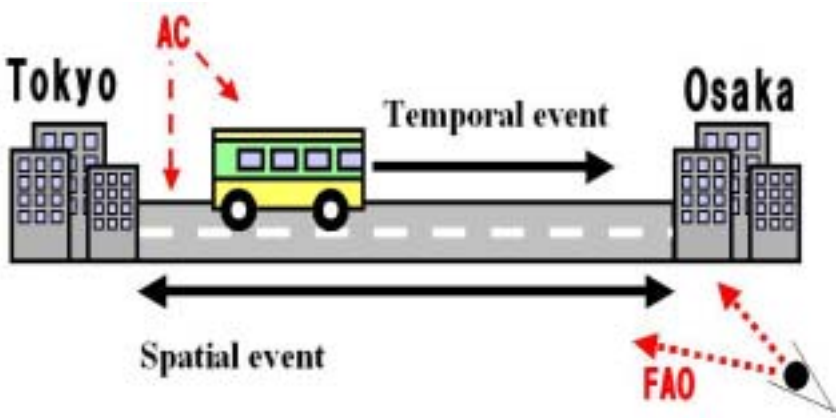
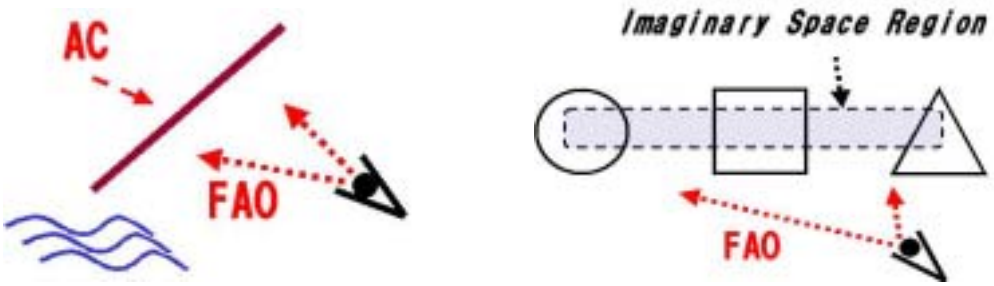


Fig.3. Event types and FAO movements.



(a) **(b)**

Fig.4. FAO movements: 'slope' (a) and 'row' (b) as spatial events.

Consequently, as shown in Fig.3, the *bus* and the FAO move together in the case of S1 while the FAO solely moves along the *road* in the case of S2. That is, ***all loci in attribute spaces correspond one to one with movements or, more generally, temporal events of the FAO***. Therefore, S3 and S4 refer to the same scene in spite of their appearances as shown in Fig.4-a where, as easily imagined, what 'sinks' or 'rises' is the FAO. The conceptual descriptions of S1 and S2 are given as (8) and (9), 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 S5 and S6 interpreted as (10) and (11), respectively. These pairs of conceptual descriptions are called ***equivalent in the PRS***, and the paired sentences are treated as ***paraphrases*** each other.

For another example of spatial event, Fig.4-b 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 S7 using the preposition 'between' and formulated as (12) or (12'), corresponding also to such concepts as 'row', 'line-up', etc. Employing ISRs and the 9-intersection model (Egenhofer, 1991; Shariff, et al, 1998), all the topological relations between two objects can be formulated in such expressions as (13) or (13') for S8, and (14) for S9, 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.

(S3) The path *sinks to* the brook.

$$(\exists x,y,p,z,k1,k2)L(x,y,p,z,A12,Gs,k1)\Pi L(x,y,\downarrow,\downarrow,A13,Gs,k2)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (8)$$

(S4) The path *rises from* the brook.

$$(\exists x,y,p,z,k1,k2)L(x,y,z,p,A12,Gs,k1)\Pi L(x,y,\uparrow,\uparrow,A13,Gs,k2)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (9)$$

(S5) Route A and Route B meet at the city.

$$(\exists x,p,y,q,k)L(x,Route\_A,p,y,A12,Gs,k)\Pi L(x,Route\_B,q,y,A12,Gs,k)\wedge city(y)\wedge p\neq q \quad (10)$$

(S6) Route A and Route B separate at the city.

$$(\exists x,p,y,q,k)L(x,Route\_A,y,p,A12,Gs,k)\Pi L(x,Route\_B,y,q,A12,Gs,k)\wedge city(y)\wedge p\neq q \quad (11)$$

(S7) □ is between Δ and ○.

$$(\exists x,y,p,q,k1,k2)(L(x,y,\Delta,\square,A12,Gs,k1)\Pi L(x,y,p,p,A13,Gs,k2))\bullet (L(x,y,\square,\circ,A12,Gs,k1)\Pi L(x,y,q,q,A13,Gs,k2))\wedge ISR(y)\wedge p=q \quad (12)$$

$$(\exists x,y,p,k1,k2)(L(x,y,\Delta,\square,A12,Gs,k1)\bullet L(x,y,\square,\circ,A12,Gs,k1))\Pi L(x,y,p,p,A13,Gs,k2)\wedge ISR(y) \quad (12')$$

(S8) Tom is in the room.

$$(\exists x,y,k)L(Tom,x,y,Tom,A12,Gs,k)\Pi L(Tom,x,In,In,A44,Gt,9IM)\wedge ISR(x)\wedge room(y) \quad (13)$$

$$(\exists x,y,k)L(Tom,x,Tom,y,A12,Gs,k)\Pi L(Tom,x,Cont,Cont,A44,Gt,9IM)\wedge ISR(x)\wedge room(y) \quad (13')$$

(S9) Tom exits the room.

$$(\exists x,y,k)L(Tom,x,y,Tom,A12,Gs,k)\Pi L(Tom,x,In,Dis,A44,Gt,9IM)\wedge ISR(x)\wedge room(y) \quad (14)$$

2.4 Attributes and standards

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. The attributes of this sense occupy the greater part of all, which implies that the sight is essential for humans to conceptualize the external world by. And this kind of classification of attributes plays a very important role in our cross-media referencing system (Hironaka & Yokota, 2004).

Table 1. Examples of attributes

	Attribute [Property]†	Linguistic expressions for attribute values.
*A01	PLACE OF EXISTE NCE [N]	He is in Tokyo. The accident happened in Osaka.
	……..	
*A11	SHAPE [N]	The cake is round.
*A12	PHYSICAL LOCATION [N]	Tom moved to Tokyo.
*A13	DIRECTION [N]	The box is to the left of the chair.
*A14	ORIENTATION [N]	The door faces to south.
*A15	TRAJECTORY [N]	The plane circled in the sky.
*A16	VELOCITY [S]	The boy runs very fast.
*A17	MILEAGE [S]	The car ran ten miles.
A18	STRENGTH OF EFFECT [S]	He is very strong.
A19	DIRECTION OF EFFECT [N]	He pulled the door.
	……	
A25	HUMIDITY [S]	The paint is still wet.
A26	VISCOSITY [S]	The liquid is oily.
A27	WEIGHT [S]	The metal is very light.
A28	TEMPERATURE [S]	It is hot today.
A29	TASTE [N]	The grapes here are very sour.
A30	ODOUR [N]	The gas is pungent.
A31	SOUND [N]	His voice is very loud.
*A32	COLOR [N]	The apple is red. Tom painted the desk white.
A33	INTERNAL SENSATION [N]	I am very tired.
A34	TIME POINT [S]	It is ten o'clock.
A35	DURATION [S]	He studies for two hours every day.
A36	NUMBER [S]	Here are many people.
A37	ORDER [S]	Tom sat next to Mary.
A38	FREQUENCY [S]	He did it twice.
A39	VITALITY [S]	The old man still alive.
	……..	

†S and N refer to 'scalar value' and 'non-scalar value', respectively. *The attributes concerning the sight.

Table 2. List of standards.

Categories of standards	Remarks
Rigid Standard	Objective standards such as denoted by measuring *units* (meter, gram, etc.).
Species Standard	The *attribute value 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*'.

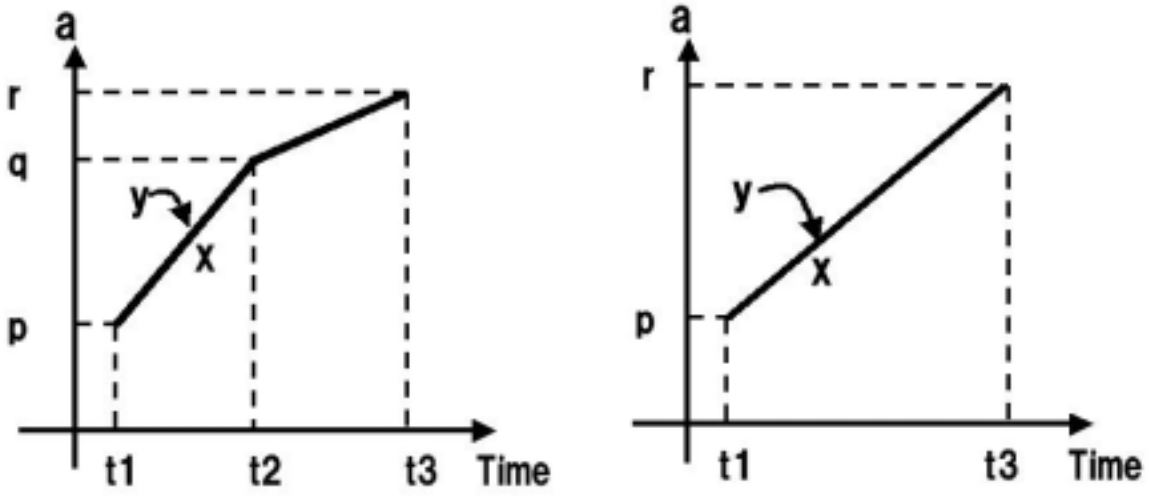
Correspondingly, six categories of standards shown in Table 2 have been extracted that are assumed necessary for representing values of each attribute in Table 1. In general, the attribute values represented by words are relative to certain standards as explained briefly in Table 2. For example, (15) and (16) are different formulations of a locus due to the different standards 'k1' and 'k2' for scaling as shown in Fig.5-a and b, 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 (17) and (18), so called, '***Postulate of Arbitrariness in Locus Articulation***'. As detailed in the next section, this postulate affects the process of conceptualization on a word based on its referents in the world.

$$(L(y,x,p,q,a,g,k1)\ \Pi\ \varepsilon(t1,t2)) \bullet (L(y,x,q,r,a,g,k1)\ \Pi\ \varepsilon(t2,t3)) \quad (15)$$

$$L(y,x,p,r,a,g,k2)\ \Pi\ \varepsilon(t1,t3) \quad (16)$$

$$(\forall p,q,r,k)\ (L(y,x,p,q,a,g,k) \bullet L(y,x,q,r,a,g,k). \supset .\ (\exists k')L(y,x,p,r,a,g,k')) \quad (17)$$

$$(\forall p,r,k)\ (L(y,x,p,r,a,g,k). \supset .\ (\exists q,k')\ L(y,x,p,q,a,g,k') \bullet L(y,x,q,r,a,g,k')) \quad (18)$$



(a) **(b)**

Fig.5. Arbitrariness in locus articulation due to standards: Standard *k1* (a) is finer than *k2* (b).

3. Mental images and word concepts

It is well known that, from the cognitive viewpoint, there are two types of mental images, namely, (a) perceptual images and (b) conceptual images. The former are live images of the current world and the latter are recalled ones, often in association with tokens such as words.

Ideally, a word concept should be associated with such a conceptual image that is abstract enough to represent the perceptual image of every matter referred to by the word. It is, however, practically impossible for an individual to obtain such a conceptual image because such instances or referents are usually too numerous for him/her to encounter and observe. In this sense, our conceptual image for a word is always imperfect or tentative to be sometimes updated by an exceptional instance just like a 'black swan'.

It is generally assumed that a word concept is an abstraction on properties and relations of the matters involved such as locations, shapes, colors, functions, potentialities, etc. In the MIDST, a word concept is to be represented as an abstract locus formula resulted from generalization on the locus formulas of a number of matters referred to by the word.

Figure 6 illustrates the mental process of conceptualization on the word 'conveyance', where a set of its referents ***Sr*** (=$\{Matter\_1,\ldots, Matter\_n\}$) are generalized by abstraction and formulated as (19). The underlined part of this formula implies that matter 'z' includes two matters 'x' and 'y' in its 'Place of existence (A01)'. As easily imagined, the variable 'z' denotes a certain referent generalized so as to represent any member of ***Sr***.

This process consists of three stages as follows. Firstly, the attributes other than 'Place of existence (A01)' and 'Physical location (A12)' are discarded. Secondly, the concrete objects 'human', 'book', etc. and their concrete attribute values are replaced by the variables 'x', 'y', 'z', 'p', 'q', etc. And finally, the relationships '≠' and '=', the most essential for this concept, are placed among these variables. The equalities (or inequalities) in 'Physical location' are determined at the precision of the standard represented by the variable 'k'.

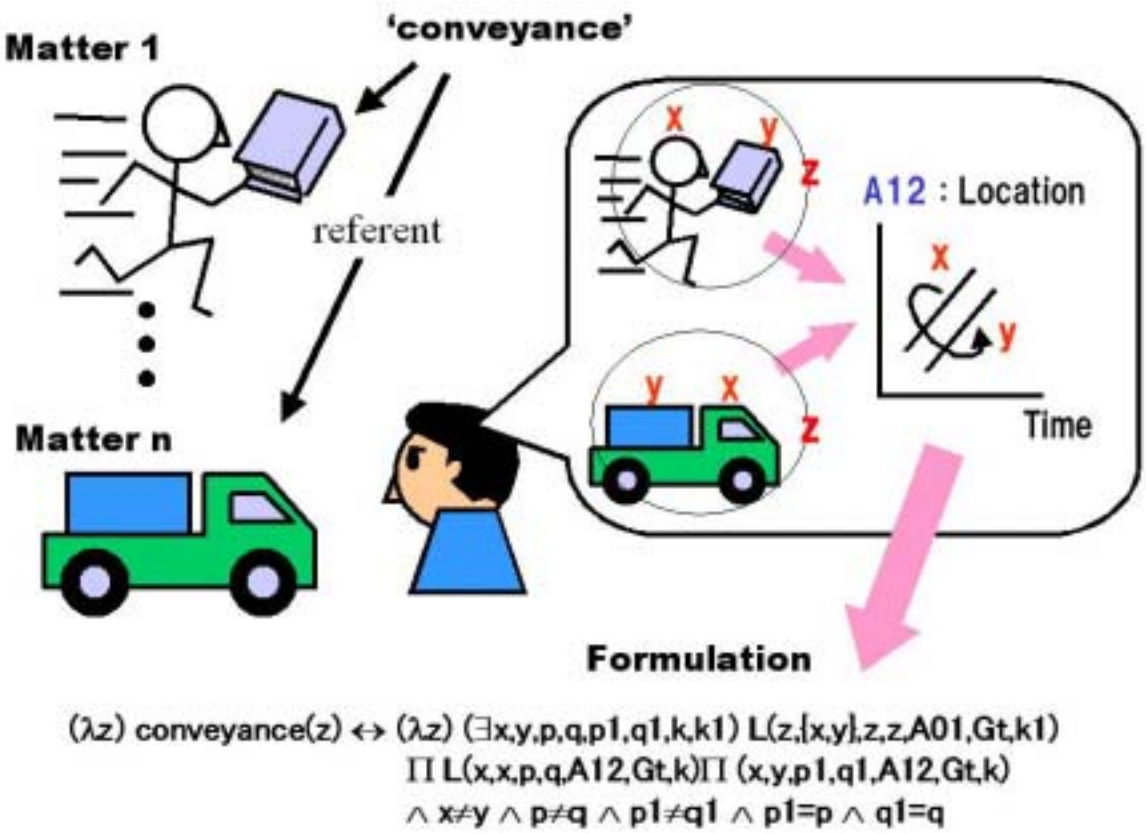


Fig.6. Conceptualization: the process of abstraction on referents of the word 'conveyance'.

$$
\begin{aligned}
&(\lambda x)\ conveyance(z) \leftrightarrow (\lambda z)(\exists x,y,p,q,p1,q1,k,k1)\ \underline{L(z,\{x,y\},z,z,A01,Gt,k1)} \\
&\qquad \Pi L(x,x,p,q,A12,Gt,k)\Pi L(x,y,p1,q1,A12,Gt,k) \\
&\qquad \wedge x\neq y \wedge p\neq q \wedge p1 \neq q1 \wedge p1=p \wedge q1=q \\
&\qquad \leftrightarrow (\lambda z)(\exists x,y,p,q,k,k1)\ \underline{L(z,\{x,y\},z,z,A01,Gt,k1)} \\
&\qquad\quad \Pi L(x,x,p,q,A12,Gt,k)\Pi L(x,y,p,q,A12,Gt,k) \wedge x\neq y \wedge p\neq q \\
&\qquad \leftrightarrow (\lambda z)(\exists x,y,p,q,k,k1)\ \underline{L(z,\{x,y\},z,z,A01,Gt,k1)} \\
&\qquad\quad \Pi L(x,\{x,y\},p,q,A12,Gt,k) \wedge x\neq y \wedge p\neq q
\end{aligned}
\qquad (19)
$$

For another example, the matter called 'snow' can be conceptualized as (20), where '_', as defined by (21), stands for the variable bound by the existential quantifier, reading 'Snow is powdered ice attracted from the sky by the earth, melts into water,…' (Hence forth, refer to Table 1 for undefined attributes.)

$$
\begin{aligned}
&(\lambda x)snow(x) \leftrightarrow (\lambda x)(\exists x1,x2,\ldots)((L(\_,x,x1,x1,A41,Gt,\_)\Pi\Pi L(Earth,x,Sky,Earth,A12,Gt,\_)) \\
&\qquad \wedge L(\_,x,x1,x2,A41,Gt,\_) \wedge powder(x1) \wedge ice(x1) \wedge water(x2) \wedge \ldots)
\end{aligned}
\qquad (20)
$$

$$L(\ldots,\_,\ldots) \leftrightarrow (\exists x)L(\ldots,x,\ldots) \quad (21)$$

For a more complicated example, the concept of 'umbrella' can be represented as (22), reading 'At raining, a human puts an umbrella in line between rain and himself/herself in order not to get wetter, …'

By the way, for the later use in Section 6, the concepts of 'rain' and 'wind' can be given as (23) and (24), reading 'Rain is water attracted from the sky by the earth, makes an object wetter, is pushed an umbrella to by a human,…,' and 'Wind is air, affects the direction of rain,… ,' respectively.

$$\begin{aligned}(\lambda x)\mathit{umbrella}(x) \leftrightarrow (\lambda x)\,(\exists x,x1,x2,y,p,q1,q2,\ldots)\,(&L(\_,x1,p,p,A13,Gt,\_)\\ &\prod((L(\_,y,x1,x,A12,Gs,\_)\bullet L(\_,y,x,x2,A12,Gs,\_))\prod L(\_,y,p,p,A13,Gs,\_)\\ &\prod L(x1,x2,q1,q2,A25,Gt,\_)\wedge ISR(y)\wedge rain(x1)\wedge human(x2)\wedge\sim(q1<q2)\ldots)\end{aligned} \quad (22)$$

$$\begin{aligned}(\lambda x)rain(x)\leftrightarrow(\lambda x)(\exists x1,x2,x3,x4,p,q,\ldots)&L(\_,x,x1,x1,A41,Gt,\_)\prod L(Earth,x,Sky,Earth,A12,Gt,\_)\\ &\prod L(x,x2,p,q,A25,Gt,\_)\prod L(x3,x4,x,x,A19,Gt,x3)\\ &\wedge water(x1)\wedge object(x2)\wedge human(x3)\wedge umbrella(x4)\wedge(p<q)\ldots\end{aligned} \quad (23)$$

$$(\lambda x)wind(x)\leftrightarrow(\lambda x)(\exists x1,x2,p,q,\ldots)L(\_,x,x1,x1,A41,Gt,\_)\wedge air(x1)\wedge(L(x,x2,p,q,A13,Gt,\_)\wedge rain(x2)\ldots \quad (24)$$

4. Word meaning descriptions

A word meaning description M_w is given by (25) as a pair of 'Concept Part (C_p)' and 'Unification Part (U_p)'.

$$M_w = [C_p{:}U_p] \quad (25)$$

The C_p of a word W is a locus formula as its concept description while its U_p is a set of operations for unifying the C_ps of W's syntactic governors or dependents. For example, the meaning of the English verb 'carry' can be given by (26).

$$[(\lambda x,y)(\exists p,q,k)\ L(x,\{x,y\},p,q,A12,Gt,k)\wedge x\neq y\wedge p\neq q:\ ARG(Dep.1,x);\ ARG(Dep.2,y);] \quad (26)$$

The U_p above consists of two operations to unify the first dependent (Dep.1) and the second dependent (Dep.2) of the current word with the variables x and y, respectively. Here, Dep.1 and Dep.2 are the 'subject' and the 'object' of 'carry', respectively. Therefore, the sentence '*Mary carries a book*' is translated into (27).

$$(\exists y,p,q,k)\ L(Mary,\{Mary,y\},p,q,A12,Gt,k)\wedge Mary\neq y\wedge p1\neq p2\wedge book(y) \quad (27)$$

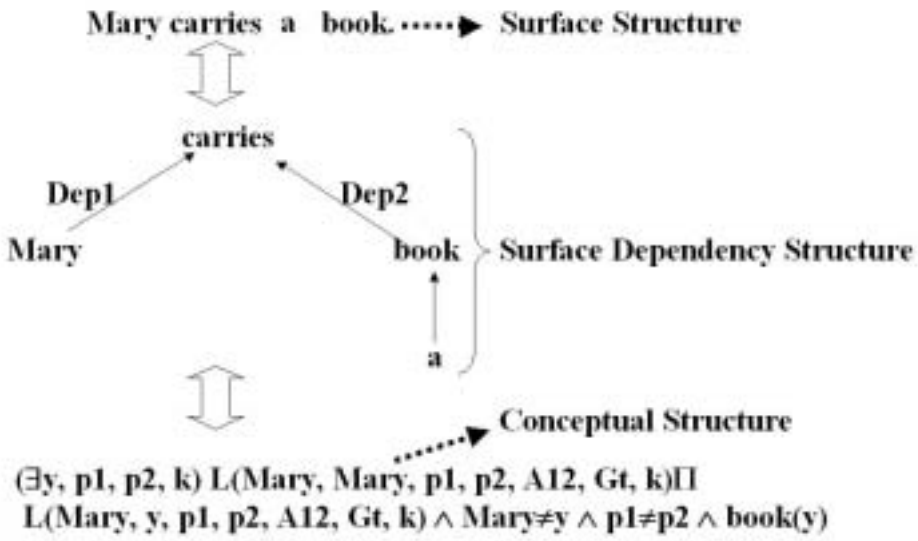


Fig.7. Process of semantic interpretation of text.

Figure 7 shows the details of the conversion process of a surface structure (text) into a conceptual structure (text meaning) through a surface dependency structure.

For another example, the meaning description of the English preposition 'through' is also given by (28).

[*(λx,y)(∃p1,z,p3,g,k,p4,k0) (L(x,y,p1,z,A12,g,k)•L(x,y,z,p3,A12,g,k))ΠL(x,y,p4,p4,A13,g,k0) ∧p1≠z ∧z≠p3: ARG(Dep.1,z); IF(Gov=Verb)→PAT(Gov,(1,1)); IF(Gov=Noun)→ARG(Gov,y);*] (28)

The U_p above is for unifying the C_ps of the very word, its governor (Gov, a verb or a noun) and its dependent (Dep.1, a noun). The second argument (1,1) of the command PAT indicates the underlined part of (28) and in general (*i*,*j*) refers to the partial formula covering from the *i*th to the *j*th atomic formula of the current C_p. This part is the pattern common to both the C_ps to be unified. This is called 'Unification Handle (U_h)' and when missing, the C_ps are to be combined simply with '∧'.

Therefore the sentences S10, S11 and S12 are interpreted as (29), (30) and (31), respectively. The underlined parts of these formulas are the results of PAT operations. The expression (32) is the C_p of the adjective 'long' implying 'there is some value greater than some standard of Length (A02),' which is often simplified as (32').

(S10) The train runs through the tunnel.

$$(\exists x,y,p1,z,p3,k,p4,k0)\ (\underline{L(x,y,p1,z,A12,Gt,k)} \bullet L(x,y,z,p3,A12,Gt,k)) \\ \Pi L(x,y,p4,p4,A13,Gt,k0) \wedge p1 \neq z \wedge z \neq p3 \wedge train(y) \wedge tunnel(z) \quad (29)$$

(S11) The path runs through the forest.

$$(\exists x,y,p1,z,p3,k,p4,k0)\ (\underline{L(x,y,p1,z,A12,Gs,k)} \bullet L(x,y,z,p3,A12,Gs,k)) \\ \Pi L(x,y,p4,p4,A13,Gs,k0) \wedge p1 \neq z \wedge z \neq p3 \wedge path(y) \wedge forest(z) \quad (30)$$

(S12) The path through the forest is long.

$$(\exists x,y,p1,z,p3,x1,k,q,k1,p4,k0)\ (L(x,y,p1,z,A12,Gs,k) \bullet L(x,y,z,p3,A12,Gs,k)) \\ \Pi L(x,y,p4,p4,A13,Gs,k0) \wedge L(x1,y,q,q,A02,Gt,k1) \wedge p1 \neq z \wedge z \neq p3 \wedge q > k1 \wedge path(y) \wedge forest(z) \quad (31)$$

$$(\exists x1,y1,q,k1) L(x1,y1,q,q,A02,Gt,k1) \wedge q > k1 \quad (32)$$

$$(\exists x1,y1,k1) L(x1,y1,Long,Long,A02,Gt,k1) \quad (32')$$

For another example, consider such somewhat complicated sentences as S13 and S14. The underlined parts are considered to refer to some events neglected in time and in space, respectively. These events are called 'Temporal Empty Event' and 'Spatial Empty Event', denoted by 'ε_t' and 'ε_s' as EEs with $g=Gt$ and $g=Gs$ at (6), respectively. The concepts of S13 and S14 are given by (33) and (34), where 'A15' and 'A17' represent the attribute 'Trajectory' and 'Mileage', respectively.

(S13) The *bus* runs 10km straight east from A to B, and <u>*after a while*</u>, at C it meets the street with the sidewalk.

$$(\exists x,y,z,p,q)\ (L(\_,x,A,B,A12,Gt,\_)\ \Pi L(\_,x,0,10km,A17,Gt,\_) \Pi L(\_,x,Point,Line,A15,Gt,\_) \Pi \\ L(\_,x,East,East,A13,Gt,\_)) \bullet \varepsilon_t \bullet (L(\_,x,p,C,A12,Gt,\_)\ \Pi L(\_,y,q,C,A12,Gs,\_) \Pi L(\_,z,y,y,A12,Gs,\_)) \\ \wedge bus(x) \wedge street(y) \wedge sidewalk(z) \wedge p \neq q \quad (33)$$

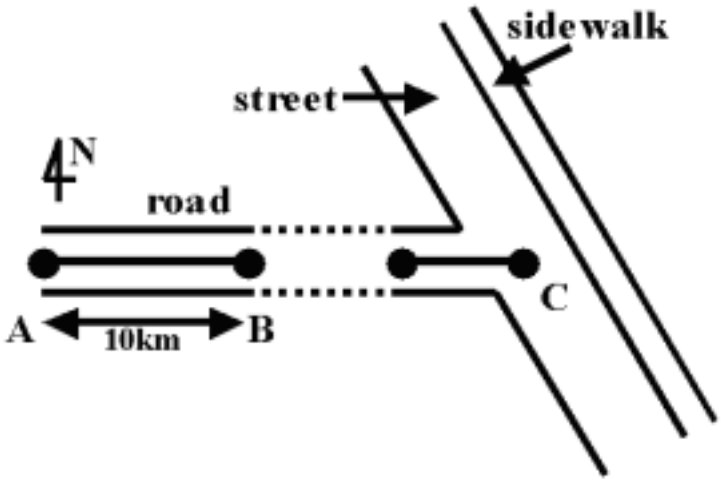


Fig.8. Pictorial interpretation of the formula (34).

(S14) The *road* runs 10km straight east from A to B, and *after a while*, at C it meets the street with the sidewalk.

$$(\exists x,y,z,p,q)(L(\_,x,A,B,A12,Gs,\_)\ \Pi\ L(\_,x,0,10km,A17,Gs,\_)\ \Pi\ L(\_,x,Point,Line,A15,Gs,\_)\Pi\ L(\_,x,East,East,A13,Gs,\_))\bullet\varepsilon_s\bullet(L(\_,x,p,C,A12,Gs,\_)\ \Pi\ L(\_,y,q,C,A12,Gs,\_)\Pi\ L(\_,z,y,y,A12,Gs,\_))\wedge road(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q \quad (34)$$

From the viewpoint of cross-media reference, the formula (34) can refer to such a spatial event depicted as the still picture in Fig.8 while (33) can be interpreted into a motion picture.

5. Fundamental semantic computation

Every version of the intelligent system IMAGES (Yokota et al, 1991; Yokota, 2005) can perform text understanding based on word meaning descriptions as follows.

Firstly, a text is parsed into a surface dependency structure (or more than one if *syntactically* ambiguous). Secondly, each surface dependency structure is translated into a conceptual structure (or more than one if *semantically* ambiguous) using word meaning descriptions. Finally, each conceptual structure is semantically evaluated.

The fundamental semantic computations on a text are to detect semantic anomalies, ambiguities and paraphrase relations.

Semantic anomaly detection is very important to cut off meaningless computations. Consider such a conceptual structure as (35), where 'A39' is the attribute 'Vitality'. This locus formula can correspond to the English sentence 'The desk is alive', which is usually semantically anomalous because a 'desk' does never have vitality in the real world projected into the attribute spaces.

$$(\exists x)L(\_,x,Alive,Alive,A39,Gt,\_)\wedge desk(x) \quad (35)$$

This kind of semantic anomaly can be detected in the following process.

Firstly, assume the concept of 'desk' as (36), where 'A29' refers to the attribute 'Taste'. The special symbols '*' and '/' are defined as (37) and (38) representing 'always' and 'no value', respectively.

$$(\lambda x)\, desk(x) \leftrightarrow (\lambda x)\,(\ldots L^*(\_,x,/,/,A29,Gt,\_) \wedge \ldots \wedge L^*(\_,x,/,/,A39,Gt,\_) \wedge \ldots) \quad (36)$$

$$X^* \leftrightarrow (\forall p,q)X\ \Pi\ \varepsilon(p,q) \quad (37)$$

$$L(\ldots,/,\ldots) \leftrightarrow \sim(\exists p)\, L(\ldots,p,\ldots) \quad (38)$$

Secondly, the postulates (39) and (40) are utilized. The formula (39) means that ***if one of two loci exists every time interval, then they can coexist***. The formula (40) states that ***a matter has never different values of an attribute at a time***.

$$X \wedge Y^* .\supset. X\ \Pi\ Y \quad (39)$$

$$L(x,y,p1,q1,a,g,k)\ \Pi\ L(z,y,p2,q2,a,g,k)\ .\ \supset.\ p1=p2 \wedge q1=q2 \quad (40)$$

Lastly, the semantic anomaly of 'alive desk' is detected by using (35)-(40). That is, the formula (41) below is finally deduced from (35)-(39) and violates the commonsense given by (40), that is, " $Alive \neq /$ ".

$$(\exists x)L(\_,x,Alive,Alive,A39,Gt,\_)\ \Pi\ L(\_,x,/,/,A39,Gt,\_) \quad (41)$$

This process above is also employed for dissolving such a syntactic ambiguity as found in S15. That is, the semantic anomaly of 'alive desk' is detected and eventually 'alive insect' is adopted as a plausible interpretation.

(S15) Look at the insect on the desk, which is still alive.

If a text has multiple plausible interpretations, it is semantically ambiguous. For example, S16 alone has two plausible interpretations (42) and (43) different at the underlined parts, implying 'Jack with the stick' and 'Tom with the stick', respectively.

(S16) Tom follows Jack with the stick.

$$(\exists x)(L(Jack,Jack,p,q,A12,Gt,\_)\underline{\Pi L(Jack,x,Jack,Jack,A12,Gt,\_)}\bullet L(Tom,Tom,p,q,A12,Gt,\_) \wedge p\neq q \wedge stick(x) \quad (42)$$

$$(\exists x)L(Jack,Jack,p,q,A12,Gt,\_)\bullet (L(Tom,Tom,p,q,A12,Gt,\_)\underline{\Pi L(Tom,x,Tom,Tom,A12,Gt,\_)}) \wedge p\neq q \wedge stick(x) \quad (43)$$

Among the fundamental semantic computations, detection of paraphrase relations is the most essential because it is for detecting equalities in semantic descriptions and the other two are for detecting inequalities in them. In our system, if two different texts are interpreted into the same locus formula, they are paraphrases of each other. For example, the sentence '*Mary goes with a book*' is interpreted into (44) which is proved to be equivalent to (27), the semantic description of '*Mary carries a book*'. In the process of this proof, the axioms (45) and (46) concerning the syntax of $\boldsymbol{L_{md}}$ are utilized.

$$(\exists y,p,q,k)L(Mary,Mary,p,q,A12,Gt,k)\Pi L(Mary,y,Mary,Mary,A12,Gt,k)\wedge p\neq q \quad (44)$$

$$(\forall x1,x2,x3,x4,p,q,a,g,k)\ L(x1,x2,p,q,a,g,k)\Pi L(x3,x4,x2,x2,a,g,k)$$

$$.\equiv. L(x1,x2,p,q,a,g,k)\Pi L(x3,x4,p,q,a,g,k) \quad (45)$$

$$(\forall x1,x2,x3,p,q,a,g,k)\ L(x1,x2,p,q,a,g,k)\Pi L(x1,x3,p,q,a,g,k) .\equiv. L(x1,\{x2,x3\},p,q,a,g,k) \quad (46)$$

6. Dialogue based on mental imagery

Viewed from the MIDST, each participant in a dialogue is to generate his/her own utterance and understand another participant's iteratively by employing his/her own mental images associated to the words involved, the environments (shared or unshared), and so on.

According to our psycholinguistic experiment on text understanding (Yokota, 1988), people often employ special ones associated with the words involved, so called, 'default' mental images that are less general but more concrete than word concepts and therefore valid only while they are consistent with the context. Consequently, people are to adjust such default images whenever they turn inconsistent with the following context.

Here is focused on such human mental image management in utterance understanding and is shown how the MIDST can simulate it by using a sample dialogue as follows:

(*Tom and Jack are friends. Tom lives in Paris while Jack, in London, with his sister Mary and her dog Pepe. Today, they are talking on telephone*.)

…
Tom: How is the weather in London?
Jack: It's raining.
Tom: Oh, it's gloomy. By the way, where is Mary now?
Jack: She's out with Pepe, making him sitting.
Tom: Oh, really? Isn't it windy?
Jack: Yes, it's blowing rather hard.
…

The following is a rough scenario of Tom's understanding of Jack's utterances above. (Refer to Section 3 for the conceptual descriptions of the words 'rain', 'umbrella' and 'wind'.)

Conceptually (i.e., using word concepts only), Tom interprets the first utterance of Jack's into (Ja1), reading 'There is rain in London.' Then, Tom infers (De1) from (Ja1) and a default image of 'rain' denoted by 'rain^{d}', employing several kinds of inference rules denoted by 'IRs' such as 'simplification (i.e., $A\wedge B.\rightarrow .A$)'. The part modified by the default image is indicated as $(\ldots)^{d}$ in (De1), implying 'Rain goes rightly downward (i.e., '$\downarrow$') and makes a house wetter.' Fig. 9-a is an example of pictorial interpretation of (De1).

$(\exists x)\ L(\_,x,London,London,A01,Gt,\_)\wedge rain(x)$ (Ja1)

Ja1, rain^{d}, IRs $\Rightarrow$ $(\exists x,x1,q1,q2)$
$(L(\_,x,\downarrow,\downarrow,A13,Gt,\_)\,I\Pi L(x,x1,q1,q2,A25,Gt,\_)\wedge house(x1)\wedge q1<q2)^{d}\wedge rain(x)$ (De1)

The second utterance of Jack's is conceptually interpreted as (Ja2), reading 'Mary and Pepe are out of the house, and she affects him so that he keeps himself in his own form of sitting (i.e., 'Sit-form' at Pepe's standard).' Then, Tom infers (De2) in use of (De1), (Ja2), rain^{d}, Mary^{d}, and IRs. The underlined part of (De2) is

significantly different from (De1), implying ‘Mary pushes an umbrella toward the rain in order not to get wetter, and Pepe is a dog.’

Furthermore, Tom infers (De3) from (De2) and umbrella[d]. The underlined part of (De3) can be depicted as Fig.9-b, reading ‘Mary is rightly under the green umbrella put between the rain and herself.’

(∃x1,y,p1,p2) L(Mary,y,{Mary,Pepe},x1,A12,Gs,_)ΠL(Mary,y,Dis,Dis,A44,Gt,9IM)
ΠL(Mary,Pepe,p1,p2,A18,Gt,_)ΠL(Pepe,Pepe,Sit-form,Sit-form,A11,Gt,Pepe)
∧ISR(y) ∧house(x1) (Ja2)

De1, Ja2, rain[d], Mary[d], IRs ⇒ *(∃x,x1,x4,y,p1,p2,q1,q2,q3,q4)*
L(Mary,y,{Mary,Pepe},x1,A12,Gs,_)ΠL(Mary,y,Dis,Dis,A44,Gt,9IM)ΠL(x,x1,q1,q2,A25,Gt,_)
ΠL(Mary,Pepe,p1,p2,A18,Gt,_)ΠL(Pepe,Pepe,Sit-form,Sit-form,A11,Gt,Pepe)
∏(L(_,x,↓,↓,A13,Gt,_)ΠL(Mary,x4,x,x,A19,Gt,Mary)∏L(x4,Mary,q3,q4,A25,Gt,_)
∧dog(Pepe)∧umbrella(x4) ∧~(q3<q4) ∧ q1<q2)[d] ∧ISR(y) ∧house(x1)∧rain(x) (De2)

De2, umbrella[d], IRs ⇒ *(∃x,x1,x4,y,y1,p1,p2,q1,q2,q3,q4)*
L(Mary,y,{Mary,Pepe},x2,A12,Gs,_)ΠL(Mary,y,Dis,Dis,A44,Gt,9IM)ΠL(x,x1,q1,q2,A25,Gt,_)
ΠL(Mary,Pepe,p1,p2,A18,Gt,_)ΠL(Pepe,Pepe,Sit-form,Sit-form,A11,Gt,Pepe)
Π(L(Mary,x4,x,x,A19,Gt,Mary)∏L(x4,Mary,q3,q4,A25,Gt,_)
∏L(_,x,↓,↓,A13,Gt,_) Π((L(Mary,y1,x,x4,A12,Gs,_)•L(Mary,y1,x4,Mary,A12,Gs,_))
∏L(Mary,y1,↓,↓,A13,Gs,_)ΠL(_,x4,Green,Green,A32,Gt,_)∧ISR(y1)∧umbrella(x4)∧~(q3<q4)∧ q1<q2)[d]
∧ISR(y) ∧house(x1)∧rain(x) (De3)

The third utterance of Jack’s is conceptually interpreted as (Ja3) reading ‘The wind affects an object hard’, and (De4) is inferred by Tom. The underlined parts of (De4) have been modified by the default image ‘wind[d]’, reading ‘The wind affects Mary hard,’ ‘The wind keeps the rain *skew* (not rightly downward (i.e., ‘\’)),’ and ‘Mary keeps also the line of the rain, the umbrella and herself *skew*,’ respectively. Figure 9-c shows a pictorial interpretation of (De4).

(∃z,x5)L(z,x5,Hard,Hard,A18,Gt,_) ∧object(x5)∧wind(z) (Ja3)

De3, Ja3, wind[d], IRs ⇒ *(∃x,x1,x4,x5,y,y1,p1,p2,q1,q2,q3,q4)*
L(Mary,y,{Mary,Pepe},x2,A12,Gs,_)ΠL(Mary,y,Dis,Dis,A44,Gt,9IM)ΠL(x,x1,q1,q2,A25,Gt,_)
ΠL(Mary,Pepe,p1,p2,A18,Gt,_)ΠL(Pepe,Pepe,Sit-form,Sit-form,A11,Gt,Pepe)Π(L(z,Mary,Hard,Hard,A18,Gt,_)
ΠL(Mary,x4,x,x,A19,Gt,Mary)∏L(x4,Mary,q3,q4,A25,Gt,_)∏L(z,x,\,\,A13,Gt,_)
((L(Mary,y1,x,x4,A12,Gs,_)•L(Mary,y1,x4,Mary,A12,Gs,_)) ∏L(Mary,y1,\,\,A13,Gs,_)
∧ISR(y1)∧umbrella(x4) ∧~(q3<q4)∧ q1<q2)[d] ∧ISR(y)∧house(x1)∧rain(x)∧wind(z) (De4)

By the way, this dialogue is to be continued as follows:

Tom: Isn't it dangerous for Mary?
Jack: No, not at all. She is in her car with Pepe seated aside.
…

(a) **(b)** **(c)**

Fig.9. Tom's default understanding of Jack's utterances: De1(a), De3(b) and De4(c).

7. Concluding remarks

The MIDST is, still under development, intended to provide a formal system represented in $\boldsymbol{L_{md}}$ for natural semantics of space and time. This formal system is one kind of applied first-order predicate logic consisting of axioms and postulates subject to human perceptive processes of space and time while the other similar systems in Artificial Intelligence (e.g., Allen, 1984; Shoham, 1989; Haddawy, 1996) are objective, namely, independent of human perception and do not necessarily keep tight correspondences with natural language.

All of the spatial prepositions have been provided with semantic descriptions in $\boldsymbol{L_{md}}$ and implemented on the intelligent system IMAGES-M. This system can perform cross-references between texts in several languages (Japanese, Chinese, Albanian and English) and pictorial patterns like maps (See APPENDIX). At our best knowledge, there is no other system that can perform cross-media operations in such a seamless way as ours (Yokota & Capi, 2005). This leads to the conclusion that our locus formula representation has made the logical expressions of event concepts remarkably computable and has proved to be very adequate to systematize cross-

media reference. This adequacy is due to its medium-freeness and its good correspondence with the performances of human sensory systems in both spatial and temporal extents. Most of computations on $\boldsymbol{L_{md}}$ are simply for unifying (or identifying) atomic loci and for evaluating arithmetic expressions such as 'p=q', and therefore we believe that our formalism can reduce the computational complexities of the traditional ones when applied to the same kinds of problems described here.

Our future work will include automatic acquisition of word concepts from sensory data and text data, and human-robot communication by natural language under real environments (Yokota, Shiraishi & Capi, 2005).

References

Allen, J.F. (1984). Towards a general theory of action and time. *Artificial Intelligence, 23-2*, 123-154.

Coventry, K. R., Prat-Sala, M. & Richards, L. V. (2001). The interplay between geometry and function in the comprehension of 'over', 'under', 'above' and 'below'. *Journal of Memory and Language*, *44,* 376-398.

Egenhofer,M. (1991). Point-set topological spatial relations. *Geographical Information Systems, 5-2*, 161-174.

Haddawy,P. (1996). A logic of time, chance, and action for repesenting plans. *Artificial Intelligence, 80-2*, 243-308.

Hironaka, D. & Yokota, M. (2004). Multimedia Description Language and Its Application to Cross-media Referencing Systems. *Proc. of IEEE DEXA workshop*, 318-323.

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

Miller,G.A. & Johnson-Laird,P.N. (1976). Language and Perception, Harvard University Press.

Shariff,A.R., 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.

Shoham,Y. (1989): Time for actions: on the relationship between time, knowledge, and action. *Proc. of IJCAI\89*, 954-959.

Sowa,J.F. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA.

Yokota,M. et al (1991). Mental-image directed semantic theory and its application to natural language understanding systems. *Proc. of NLPRS'91*, 280-287.

Yokota,M. (1988). A psychological experiment on human understanding process of natural language. *Trans. of IEICE Japan*, J71D-10, 2120-2127.

Yokota,M., Shiraishi,M., Capi,G. (2005). Human-robot communication through a mind model based on the Mental Image Directed Semantic Theory. *Proc. of the 10th International Symposium on Artificial Life and Robotics (AROB '05)*, 695-698.

Yokota,M. (2005). An approach to natural language understanding based on a mental image model. *Proc. of the 2nd International Workshop on Natural Language Understanding and Cognitive Science*, 22-31.

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.

APPENDIX

The intelligent system IMAGES-M, one kind of expert system in Artificial Intelligence, consists of Knowledge Base (KB), Inference Engine (IE) and five kinds of user interfaces (TPU, PPU, SDPU, ADPU and SPU) as shown in Fig.10. Figures 11 and 12 show some examples of cross-media dialogues between humans and IMAGES-M.

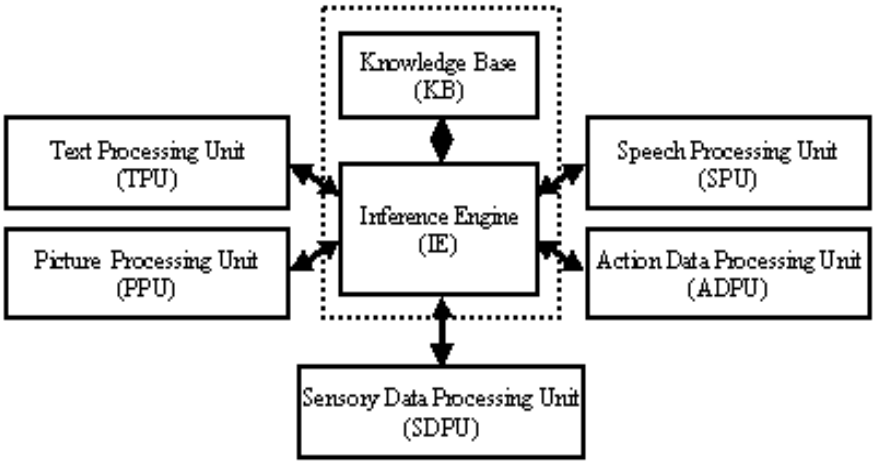
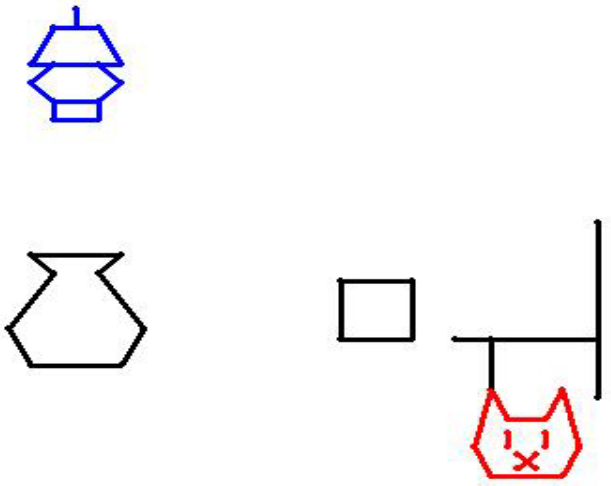


Fig.10. Configuration of the intelligent system IMAGES-M.

Fig.11. A real map that IMAGES-M generated from its corresponding locus formula.

the chair is 3m to the right of the big pot.
猫は椅子の1m下にいる
Macja eshte e kuqe.
the small box is 1m to the left of the chair.
the big blue lamp is 2m above the pot.

(a)



(b)

H: **?猫 是 紅的** (Is the cat red ?)
S : **是** (yes)
H: **?何が椅子と花瓶の間にある**
(What is between the chair and the flower-pot ?)
S: **箱** (box)
H: **Is the box between the cat and the pot?**
S: **NO**
H: **Eshte kutia midis maces dhe llampes?**
(Is the box between the cat and the lamp?)
S: **PO** (yes)

(c)

Fig.12. Dialogues between Humans and IMAGES-M: Human's utterance (a), its pictorial understanding by IMAGES-M (b), and Q-A about the picture (H: human, S: IMAGES-M) (c).