18

Towards Artificial Communication Partners with a Multiagent Mind Model Based on Mental Image Directed Semantic Theory

Masao Yokota
Fukuoka Institute of Technology
Japan

1. Introduction

In recent years, there have been developed various types of real or virtual robots as artificial communication partners. However, they are to play their roles according to programmed actions to stimuli and have not yet come to understand or imitate delicate mental functions of their human partners such as *Kansei*, one of the topics in this chapter. Kansei evaluates non-scientific matters such as art, music, natural scenery, etc. by Kansei words (e.g., 'heart-calming', 'fantastic', 'grotesque') (Fukuda et al, 1998; Sugita et al, 2003; Butz et al, 2005) and 'Artificial Kansei' namely 'Kansei for robots' is expected to play a part in 'artificial or robotic individuality' (Tauchi et al, 2006).

The author has proposed a human mind model consisting of Stimulus, Knowledge, Emotion and Response Processing Agents, intending the intelligent system IMAGES-M (Yokota, 2005a; Shiraishi et al, 2005) to understand and imitate miscellaneous human mental functions involving *emotion processing* as well as *knowledge processing* originally aimed at. For example, Kansei is defined and realized as tight collaboration of Knowledge and Emotion Processing Agents. This multiagent mind model is much simpler than Minsky's (Minsky, 1986) and its most remarkable feature is that the agents involved communicate with one another by exchanging and computing mental images represented in the formal language L_{md} developed for integrated representation and computation of multimedia information (Yokota, 2005, 2006) while other multiagent systems were designed to employ special programming languages for inter-agent communication (e.g., Labrou et al, 1999; Vieira et al, 2007).

IMAGES-M is originally intended for integrated multimedia understanding as *knowledge processing* for intuitive human-robot interaction such that may happen between ordinary (or non-expert) people and home robots (Yokota et al, 2008). Here, 'integrated multimedia understanding' means especially 'multimedia understanding based on such a knowledge representation common to multiple information media (i.e., natural language, picture, music, gesture, etc.) that can facilitate cross-media operations (Yokota et al, 1984; Eakins & Graham, 1999; Kherfi et al, 2004; Yokota & Capi, 2005a)'. For ordinary people, however, natural language is the most important because it can convey the exact intention of the sender to the receiver due to its syntax and semantics common to its users, which is not necessarily the case for another information medium such as gesture or so. Therefore, natural language can as well play the most crucial role in intuitive human-robot interaction and actually IMAGES-M is a language-centered multimedia understanding system where its attention mechanism is to be controlled efficiently in a top-down way according to people's suggestions in natural language (Yokota, 2007).

For integrated multimedia understanding intended here, it is essential to develop a systematically computable knowledge representation language (KRL) as well as representation-free technologies (Brooks, 1986) such as neural networks for unstructured sensory/motory data processing (i.e., *stimulus/response processing*). This type of language is indispensable to *knowledge-based* processing such as *understanding* sensory events, *planning* appropriate actions and *knowledgeable* communication with ordinary people in natural language, and therefore it needs to have at least a good capability of representing spatiotemporal events that correspond to humans'/robots' sensations and actions in the real world (Yokota & Capi, 2005b).

Most of conventional methods have provided robotic systems with such quasi-natural language expressions as 'move(*Velocity*, *Distance*, *Direction*)', 'find(*Object*, *Shape*, *Color*)', etc. for human instruction or suggestion, uniquely related to computer programs for deploying sensors/motors as their semantics (e.g., Coradeschi & Saffiotti, 2003; Drumwright et al, 2006). These expression schemas, however, are too linguistic or coarse to represent and compute sensory/motory events in such an integrated way as intuitive human-robot interaction intended here. This is also the case for AI planning ("action planning") which deals with the development of representation languages for planning problems and with the development of algorithms for plan construction (Wilkins & Myers, 1995).

In order to solve this problem, the author has employed the formal language so called 'Language for Mental-image Description ($\boldsymbol{L_{md}}$)' proposed in his original semantic theory 'Mental Image Directed Semantic Theory (MIDST)' (e.g., Yokota, 2005, 2006), the main topic in this chapter.

MIDST is based on the hypothesis that natural language understanding in humans is basically mental image processing and concerns:

(i) Modeling of the human mind as a society of agents;
(ii) Modeling of omnisensory mental image in humans;
(iii) Modeling of conceptualization in humans;
(iv) Designing the formal language $\boldsymbol{L_{md}}$ for mental image description;
(v) Formulation of word concepts as mental images in $\boldsymbol{L_{md}}$;
(vi) Mutual translation between expressions in miscellaneous information media (e.g., natural language, picture, robotic action, etc.) and those in $\boldsymbol{L_{md}}$;
(vii) Computation on $\boldsymbol{L_{md}}$ expressions;
(viii) Formalization of human mental competence and performance as a deductive system in $\boldsymbol{L_{md}}$;

The final goal of MIDST is to realize artificial communication partners with a good capability of intuitive interaction with ordinary people and the items (i)-(viii) above are assumed to be its subgoals. The key idea of MIDST is the model of human attention-guided (i.e., active) perception yielding omnisensory images that inevitably reflect certain movements of the focus of attention of the observer (FAO) scanning certain matters in the world, either inside or outside of the mind. More analytically, these omnisensory images are associated with spatiotemporal changes (or constancies) in certain attributes of the matters scanned by FAO and modeled as temporally parameterized "loci in attribute spaces", so called, to be formulated in the formal language L_{md}. This language has already been implemented on several types of computerized intelligent systems including IMAGES-M (e.g., Yokota et al, 1984; Oda, et al, 2001; Amano, et al, 2005; Yokota & Capi, 2005a). The most remarkable feature of L_{md} is its capability of formalizing spatiotemporal matter concepts grounded in human/robotic sensation while the other similar KRLs are designed to describe the logical relations among conceptual primitives represented by lexical tokens (e.g., Dorr & Bonnie, 1997; Zarri, 1997; Sowa, 2000). Moreover, in L_{md} expression are hinted what and how should be attended to in the world as analogy of human FAO movement and thereby the robotic attention can be controlled in a top-down way (Yokota, 2007), which is the author's answer to the essential issue in robotic imitation, namely, how to control robotic attention mechanism efficiently (e.g., Demiris & Khadhouri, 2006).

The remainder of this chapter is organized as follows. Section 2 presents MIDST, focusing on the multiagent mind model, the omnisensory mental image model and the formal language L_{md} with liguistic or pictorial manifestations for its validation. Section 3 details about grounding natural language expressions in mental images in view of natural language processing by computers. In Section 4, the mental fucntion Kansei is modeled as collaboration of **Kn** and **Em** viewed from artificial or robotic individuality. Section 5 presents a discussion on applications and further developments for the language L_{md} presented in this chapter. Conclusions and planned future work are given in the final section.

2. Mental Image Directed Semantic Theory

2.1 Multiagent mind model

Figure 1 shows the multiagent mind model proposed here, consisting of Stimulus, Knowledge, Emotion and Response Processing Agents. This is a functional model of human central nervous system consisting of the brain and the spine. These agents are to communicate with one another by exchanging and computing mental images represented in the formal language L_{md}. Their basic performances are as follows.

1) **Stimulus Processing Agent** (**St**) receives stimuli from the world (**W**) and encodes them into mental images (i.e. encoded sensations) such as *"I sensed something oily."* (if verbalized in English.)
2) **Knowledge Processing Agent** (**Kn**) evaluates mental images received from the other agents based on its memory (e.g. knowledge), producing other mental images such as *"It is false that the earth is flat."*
3) **Emotion Processing Agent** (**Em**) evaluates mental images received from the other agents based on its memory (e.g. instincts), producing other mental images such as *"I like the food."*
4) **Response Processing Agent** (**Re**) converts mental images (i.e. encoded actions such as *"I'll walk slowly."*) received from the other agents into real actions against **W**.

A *performance* $\mathbf{P}$ against a *stimulus* X with a *result* Y at each agent can be formalized as a function by the expression (1).

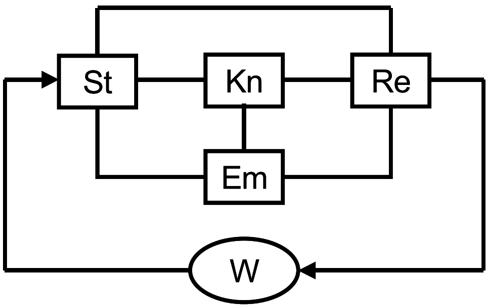
$$Y=\mathbf{P}(X) \tag{1}$$

where

P: a combination of ***Atomic Performances*** defined later in association with ***Attribute Spaces***;
X: a spatiotemporal distribution of stimuli from **W** to **St** or a mental image for another agent.
Y: a series of signals to drive actuators for **Re** or a mental image for another agent.

For example, all the agents are to work during understanding information media such as natural language, picture, music, gesture, etc., sometimes performing Kansei by tight collaboration of **Kn** and **Em** as detailed later while **St** and **Re** are exclusively to work during *reflection* so called.

A performance **P** is assumed as a function formed either consciously or unconsciously, or in other words, either with or without reasoning. In a conscious case, a set of atomic performances are to be chosen and combined according to *X* by a meta-function, so called, '***Performance Selector***' assumed as '***Conscience***'. On the contrary, in an unconscious case, such a performance as is associated most strongly with *X* is to be applied automatically as is in the case of reflection.



Multiagent model of human mind (St: Stimulus Processing Agent; Kn: Knowledge Processing Agent; Em: Emotion Processing Agent; Re: Response Processing Agent; W: World surrounding human mind, including his/her body).

2.2 Omnisensory mental image model and formal language L_{md}

Here are described the omnisensory mental image model and the syntax and semantics of L_{md} in association with the mental image model. In MIDST, word meanings are defined in association with mental images, not limited to visual but omnisensory, modeled as "Loci in Attribute Spaces" so called. See Fig.2-a and assume that the human is observing the phenomenon where the triangular gray object is moving in the sinusoidal trajectory and that its corresponding sensations (i.e., sensory images) are being caused in his/her mind. In this case, the moving triangular gray object is assumed to be perceived as the loci in the three attribute spaces, namely, those of 'Location', 'Color' and 'Shape' in the observer's mind. As easily imagined, attribute spaces correspond with human sensory systems and the loci represent certain sensations of the phenomena outside or inside human minds. From the viewpoint of artifact, an attribute space stands for a certain measuring instrument or sensor just like a chronograph, barometer, thermometer or so and the loci represent the movements of its indicator. The performance of an attribute space is the model of '***Atomic Performance***' introduced in Section 2.1.

These loci are to be articulated by "Atomic Locus" over a certain absolute time interval [t_i, t_f] as depicted in Fig.2-b and formulated as (2) in L_{md}, where the interval is suppressed because people are not aware of absolute time (nor always consult a chronograph).

$$L(x,y,p,q,a,g,k) \quad (2)$$

The expression (2) works as a formula in many-sorted predicate logic, where "L" is a predicate constant with five types of terms: "Matter" (at 'x' and 'y'), "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 (3) and (4)) to appear at Values or Standard in order to represent their values in each place at the time or over the time-interval. Moreover, when it is not so significant to discern ECs or Standards, anonymous variables, usually symbolized as '_', can be employed in their places (See (23) for example). A logical combination of atomic locus formulas defined as a well-formed formula (i.e., wff) in predicate logic is called simply 'Locus Formula'.

The intuitive interpretation of (2) 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 certain absolute time-interval, where the values 'p' and 'q' are relative to the standard 'k'."

In (2), when $g=G_t$, the locus indicates monotonic change (or constancy) of the attribute in time domain, and when $g=G_s$, that in space domain. 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 (3) and (4), respectively, where 'A_{12}' denotes the attribute 'Physical Location'. These two formulas are different only at the term 'Event Type'.

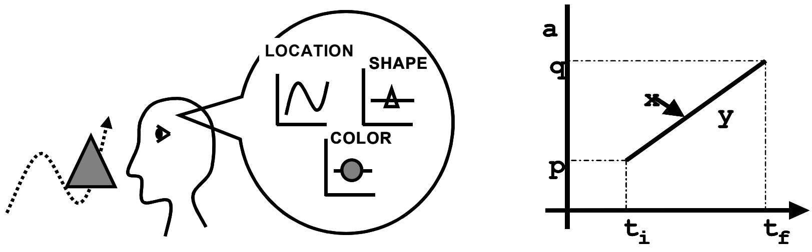
(S1) The bus runs from Tokyo to Osaka.

(S2) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A_{12},G_t,k)\wedge bus(y) \quad (3)$$

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A_{12},G_s,k)\wedge road(y) \quad (4)$$

It has been often argued that human active sensing processes may affect perception and in turn conceptualization and recognition of the physical world while such cognitive processes or products have seldom been formulated for computation (e.g., Leisi, 1961; Noton, 1970; Gardenfors, 2000; Langacker, 2005). The author has hypothesized that 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, it is hypothesized that 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.3, the *bus* and FAO move together in the case of S1 while FAO solely moves along the *road* in the case of S2. That is, ***all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal events of FAO***.



(a) (b)

Fig. 2. Mental image model (a) and Atomic Locus in Attribute Space (b)

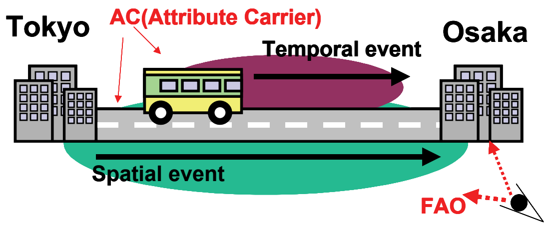


Fig. 3. FAO movements and Event types

2.3 Tempological connectives

The duration of a locus corresponds to an absolute time-interval over which FAO is put on the corresponding phenomenon outside or inside the mind. Such an absolute time-interval is suppressed in an atomic locus formula because it is assumed that people cannot measure the absolute time by any chronograph but a certain relative time (Actually, people do not always consult a chronograph even if they can). MIDST has employed 'tempo-logical connectives (TLCs)' denoting both logical and temporal relations between loci by themselves because these must be considered simultaneously in locus articulation.

A tempo-logical connective K_i is defined by (5), where τ_i, χ and K refer to one of the temporal relations indexed by an integer 'i', a locus, and an ordinary binary logical connective such as the conjunction '$\wedge$', respectively. The definition of τ_i is given in Table 1 from which the theorem (6) can be deduced. This table shows the complete list of topological relations between two intervals, where 13 types of relations are discriminated by τ_i ($-6 \leq$ i ≤ 6). This is in accordance with Allen's notation (Allen, 1984), which, to be strict, is exclusively for 'temporal conjunctions ($=\wedge_i$)' as introduced below.

$$\chi_1 \, K_i \, \chi_2 \Leftrightarrow (\chi_1 \, K \, \chi_2) \wedge \tau_i(\chi_1, \chi_2) \tag{5}$$

$$\tau_{-i}(\chi_2, \chi_1) \equiv \tau_i(\chi_1, \chi_2) \quad (\forall i \in \{0, \pm1, \pm2, \pm3, \pm4, \pm5, \pm6\}) \tag{6}$$

The TLCs used most frequently are 'SAND ($\wedge_0$)' and 'CAND ($\wedge_1$)', standing for 'Simultaneous AND' and 'Consecutive AND' and conventionally symbolized as 'Π' and '•', respectively. For example, the concepts of the English verbs 'carry' and 'return' are to be defined as (7) and (8), respectively. These can be depicted as Fig.4-a and b, respectively. The expression (9) is the definition of the English verb concept 'fetch' depicted as Fig.4-c. This implies such a temporal event that 'x' goes for 'y' and then comes back with it. In the same way, the English verb concept 'hand' or 'receive' depicted as Fig.4-d is defined equivalently as (10) or its abbreviation (10') where ECs are merged into a set.

$$(\lambda x,y)\mathrm{carry}(x,y)\Leftrightarrow(\lambda x,y)(\exists p,q,k)L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q \tag{7}$$

$$(\lambda x)\mathrm{return}(x)\Leftrightarrow(\lambda x)(\exists p,q,k)L(x,x,p,q,A_{12},G_t,k)\bullet L(x,x,p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q \tag{8}$$

$$(\lambda x,y)\mathrm{fetch}(x,y)\Leftrightarrow(\lambda x,y)\ (\exists p_1,p_2,k)\ L(x,x,p_1,p_2,A_{12},G_t,k)\bullet$$

$$((\underline{L(x,x,p_2,p_1,A_{12},G_t,k)\Pi L(x,y,p_2,p_1,A_{12},G_t,k)})\wedge x\neq y\wedge p_1\neq p_2 \tag{9}$$

$$(\lambda x,y,z)\mathrm{hand}(x,y,z).\equiv.(\lambda x,y,z)\mathrm{receive}(z,,y,x)$$

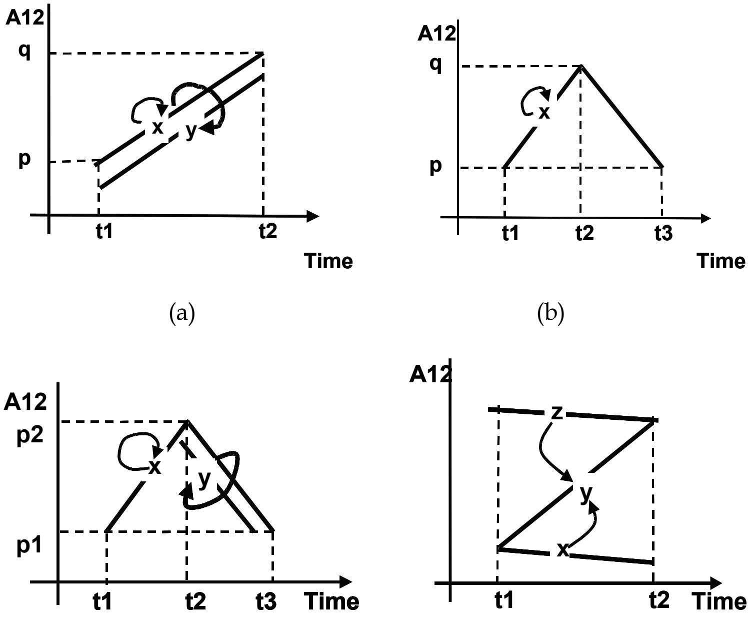
$$\Leftrightarrow(\lambda x,y,z)(\exists k)L(x,y,x,z,A_{12},G_t,k)\Pi L(z,y,x,z,A_{12},G_t,k)\wedge x\neq y\wedge y\neq z\wedge z\neq x \tag{10}$$

$$(.\equiv.\ (\lambda x,y,z)(\exists k)L(\{x,z\},y,x,z,A_{12},G_t,k)\wedge x\neq y\wedge y\neq z\wedge z\neq x) \tag{10'}$$

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 (A_{12})', for example, *start*, *stop*, *meet*, *separate*, *carry*, *return*, etc.

Temporal relations and definition of τ_i†			Allen's notation
χ_1 +·····························+ χ_2 +·····························+	$t_{11}=t_{21}$ $\wedge t_{12}=t_{22}$	$\tau_0(\chi_1, \chi_2)$	equals(χ_1,χ_2)
		$\tau_0(\chi_2, \chi_1)$	equals(χ_2, χ_1)
χ_1 +·············+ χ_2 +··············+	$t_{12}=t_{21}$	$\tau_1(\chi_1, \chi_2)$	meets(χ_1, χ_2)
		$\tau_{-1}(\chi_2, \chi_1)$	met-by(χ_2, χ_1)
χ_1 +·············+ χ_2 +·····························+	$t_{11}=t_{21}$ $\wedge t_{12}<t_{22}$	$\tau_2(\chi_1, \chi_2)$	starts(χ_1, χ_2)
		$\tau_{-2}(\chi_2, \chi_1)$	started-by(χ_2, χ_1)
χ_1 +·············+ χ_2 +·····························+	$t_{11}>t_{21}$ $\wedge t_{12}<t_{22}$	$\tau_3(\chi_1, \chi_2)$	during(χ_1, χ_2)
		$\tau_{-3}(\chi_2, \chi_1)$	contains(χ_2, χ_1)
χ_1 +·············+ χ_2 +·····························+	$t_{11}>t_{21}$ $\wedge t_{12}=t_{22}$	$\tau_4(\chi_1, \chi_2)$	finishes(χ_1, χ_2)
		$\tau_{-4}(\chi_2, \chi_1)$	finished-by(χ_2, χ_1)
χ_1 +·······+ χ_2 +··········+	$t_{12}<t_{21}$	$\tau_5(\chi_1, \chi_2)$	before(χ_1, χ_2)
		$\tau_{-5}(\chi_2, \chi_1)$	after(χ_2, χ_1)
χ_1 +··················+ χ_2 +···················+	$t_{11}<t_{21}\wedge t_{21}<t_{12}$ $\wedge t_{12}<t_{22}$	$\tau_6(\chi_1, \chi_2)$	overlaps(χ_1, χ_2)
		$\tau_{-6}(\chi_2, \chi_1)$	overlapped-by(χ_2, χ_1)

Table 1. List of temporal relations (†χ_1 and χ_2 exist during [t_{11}, t_{12}] and [t_{21}, t_{22}], respectively)



(a) (b)

(c) (d)

Fig. 4. Depiction of loci: ‘carry’ (a), ‘return’ (b), ‘fetch’ (c) and ‘hand/receive’ (d)

In order for explicit indication of time points, a very important concept called ‘Empty Event (EE)’ denoted by ‘ε’ is introduced. An EE stands only for absolute time elapsing and is explicitly defined as (11) with the attribute ‘Time Point (A_{34})’ and the standard ‘K_{Ta}’ denoting *absolute time*, where t_i and t_j are conventionally given as real numbers with the condition $t_i<t_j$. According to this scheme, the duration $[t_a, t_b]$ of an arbitrary locus χ can be expressed as (12).

$$\varepsilon([t_i,t_j]) \Leftrightarrow (\exists x,y,g)\ L(x,y,\ t_i,t_j,A_{34},g,K_{Ta}) \tag{11}$$

$$\chi \Pi \varepsilon([t_a,t_b]) \tag{12}$$

Any pair of loci temporally related in certain attribute spaces can be formulated as (13)-(17) in exclusive use of SANDs, CANDs and EEs. For example, the loci shown in Fig.5-a and b correspond to the formulas (14) and (17), respectively.

$$\chi_1 \wedge_2 \chi_2 \ .\equiv.\ (\chi_1 \bullet \varepsilon) \Pi \chi_2 \tag{13}$$

$$\chi_1 \wedge_3 \chi_2 \ .\equiv.\ (\varepsilon_1 \bullet \chi_1 \bullet \varepsilon_2) \Pi \chi_2 \tag{14}$$

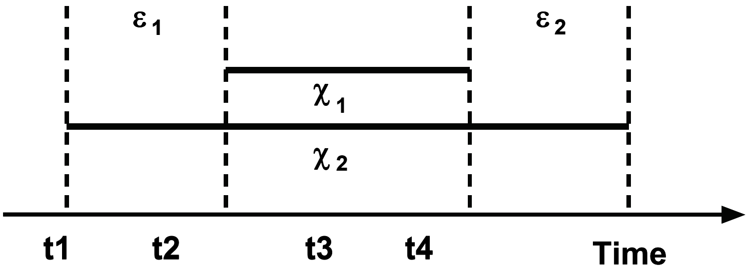
$$\chi_1 \wedge_4 \chi_2 \ .\equiv.\ (\varepsilon \bullet \chi_1) \Pi \chi_2 \tag{15}$$

$$\chi_1 \wedge_5 \chi_2 .\equiv. \chi_1 \bullet \varepsilon \bullet \chi_2 \tag{16}$$

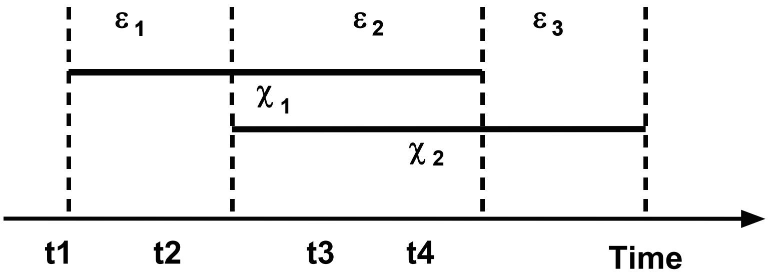
$$\chi_1 \wedge_6 \chi_2 .\equiv. (\chi_1 \bullet \varepsilon_3)\Pi(\varepsilon_1 \bullet \chi_2)\Pi(\varepsilon_1 \bullet \varepsilon_2 \bullet \varepsilon_3) \tag{17}$$

Employing TLCs, tempo-logical relationships between miscellaneous event concepts can be formulated without explicit indication of time intervals. For example, an event 'fetch(x,y)' is necessarily *finished by* an event 'carry(x,y)' as indicated by the underline at (9). This fact can be formulated as (18), where '$\supset_{-4}$' is the '*implication* ($\supset$)' furnished with the temporal relation '*finished-by* (τ_{-4})'. This kind of formula is not an axiom but a theorem deducible from the definitions of event concepts in the deductive system intended here.

$$(\forall x,y)(\text{fetch}(x,y) \supset_{-4} \text{carry}(x,y)) \tag{18}$$



(a)



(b)

Fig. 5. Tempological relations: (a) during(χ_1, χ_2) and (b) overlaps(χ_1, χ_2)

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 as shown in Table 2 exclusively from Japanese and English words (e.g., Roget, 1975). 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 2, 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 operating system (Yokota & Capi, 2005a).

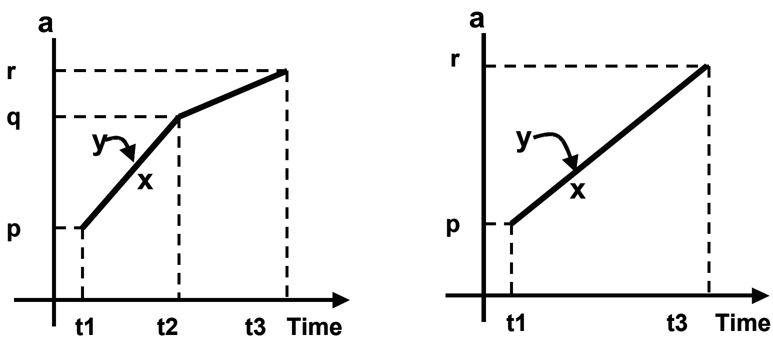
Correspondingly, six categories of standards shown in Table 3 have been extracted after the conventional categorization (Leisi, 1961) that are assumed necessary for representing values of each attribute in Table 2. In general, the attribute values represented by words are relative to certain standards as explained briefly in Table 3. For example, (19) and (20) are different formulations of a locus due to the different standards 'k1' and 'k2' for scaling as shown in Fig.6-a and b, respectively. That is, whether the point (t_2, 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 (21) and (22), 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.

Code	Attribute [Property]†	Linguistic expressions for attribute values
*A_{01}	PLACE OF EXISTE NCE [N]	The accident happened in Osaka.
*A_{02}	LENGTH [S]	The stick is 2 meters long.
	…………………………….	
*A_{11}	SHAPE [N]	The cake is round.
*A_{12}	PHYSICAL LOCATION [N]	Tom moved to Tokyo.
*A_{13}	DIRECTION [N]	The box is to the left of the chair.
*A_{14}	ORIENTATION [N]	The door faces to south.
*A_{15}	TRAJECTORY [N]	The plane circled in the sky.
*A_{16}	VELOCITY [S]	The boy runs very fast.
*A_{17}	MILEAGE [S]	The car ran ten miles.
A_{18}	STRENGTH OF EFFECT [S]	He is very strong.
A_{19}	DIRECTION OF EFFECT [N]	He pulled the door.
	……………………………	
A_{28}	TEMPERATURE [S]	It is hot today.
A_{29}	TASTE [N]	The grapes here are very sour.
A_{30}	ODOUR [N]	The gas is pungent.
A_{31}	SOUND [N]	His voice is very loud.
*A_{32}	COLOR [N]	Tom painted the desk white.
A_{33}	INTERNAL SENSATION [N]	I am very tired.
A_{34}	TIME POINT [S]	It is ten o'clock.
A_{35}	DURATION [S]	He studies for two hours every day.
A_{36}	NUMBER [S]	Here are many people.
A_{37}	ORDER [S]	Tom sat next to Mary.
A_{38}	FREQUENCY [S]	He did it twice.
A_{39}	VITALITY [S]	The old man still alive.
*A_{44}	TOPOLOGY [N]	He is in the room.
*A_{45}	ANGULARITY [S]	The knife is dull.

Table 2. Examples of attributes (†S: 'scalar value', N : 'non-scalar value')

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	Tom is taller *than Jim*. The origin of an order such as 'next' must be declared explicitly just as 'next *to him*'.

Table 3. List of standards



(a) (b)

Fig. 6. Arbitrariness in locus articulation due to standards: Standard k_1 (a) is finer than k_2 (b)

$$(L(y,x,p,q,a,g,k_1)\Pi\ \varepsilon([t_1,t_2]))\bullet(L(y,x,q,r,a,g,k_1)\Pi\ \varepsilon([t_2,t_3])) \tag{19}$$

$$L(y,x,p,r,a,g,k_2)\Pi\ \varepsilon([t_1,t_3]) \tag{20}$$

$$(\forall p,q,r,k)(L(y,x,p,q,a,g,k)\bullet L(y,x,q,r,a,g,k).\supset.(\exists k')L(y,x,p,r,a,g,k')\wedge k'\neq k) \tag{21}$$

$$(\forall p,r,k)(L(y,x,p,r,a,g,k).\supset.(\exists q,k')\ L(y,x,p,q,a,g,k')\bullet L(y,x,q,r,a,g,k')\wedge\ k'\neq k) \tag{22}$$

3. Mind model and natural language

3.1 Spatiotemporal expressions and perceptual processes

As already mentioned in Section 2, all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal events of FAO. Therefore, an event expressed 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. And this is why S3 and S4 can refer to the same scene in spite of their appearances, where what 'sinks' or 'rises' is FAO as illustrated in Fig.7-a and whose conceptual descriptions are given as (23) and (24), respectively, where 'A_{13}', '↑' and '↓' refer to the attribute 'Direction' and its values 'upward' and 'downward', respectively.

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

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

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 as interpreted approximately into (25) and (26), respectively. These pairs of conceptual descriptions are called equivalent in the PRS, and the paired sentences are treated as paraphrases each other.

(a) (b)

Fig. 7. FAO movements: 'slope' (a) and 'row' (b) as spatial events

$$(\exists y,p,z)L(\_,y,p,z,A_{12},G_s,\_)\Pi L(\_,y,\downarrow,\downarrow,A_{13},G_s,\_)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (23)$$

$$(\exists y,p,z)L(\_,y,z,p,A12,G_s,\_)\Pi L(\_,y,\uparrow,\uparrow,A_{13},G_s,\_)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (24)$$

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

$$(\exists p,y,q)L(\_,Route\_A,p,y,A_{12},G_s,\_)\Pi L(\_,Route\_B,q,y,A_{12},G_s,\_)\wedge city(y)\wedge p\neq q \quad (25)$$

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

$$(\exists p,y,q)L(\_,Route\_A,y,p,A_{12},G_s,\_)\Pi L(\_,Route\_B,y,q,A_{12},G_s,\_)\wedge city(y)\wedge p\neq q \quad (26)$$

For another example of spatial event, Fig.7-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 (27) or (27'), corresponding also to such concepts as 'row', 'line-up', etc. Employing ISRs and the 9-intersection model (Egenhofer, 1991), all the topological relations between two objects can be formulated in such expressions as (28) or (28') for S8, and (29) for S9, where '*In*', '*Cont*' and '*Dis*' are the values 'inside', 'contains' and 'disjoint' of the attribute 'Topology (A_{44})' with the standard '9-intersection model (K_{9IM})', respectively. Practically, these topological values are given as 3×3 matrices with each element equal to 0 or 1 and therefore, for example, '*In*' and '*Cont*' are transpositional matrices each other.

(S7) The square is between the triangle and the circle.
(S8) Tom is in the room.
(S9) Tom exits the room.

$$(\exists y,x_1,x_2,x_3,p,q)\ (L(\_,y,x_1,x_2,A_{12},G_s,\_)\Pi L(\_,y,p,p,A_{13},G_s,\_))\bullet(L(\_,y,x_2,x_3,A_{12},G_s,\_)$$
$$\Pi L(\_,y,q,q,A_{13},G_s,\_))\wedge ISR(y)\wedge p=q\ \wedge triangle(x_1)\wedge square(x_2)\ \wedge circle(x_3) \quad (27)$$

$$(\exists y,x_1,x_2,x_3,p)(L(\_,y,x_1,x_2,A_{12},G_s,\_)\bullet L(\_,y,x_2,x_3,A_{12},G_s,\_))\Pi L(\_,y,p,p,A_{13},G_s,\_)$$
$$\wedge ISR(y)\ \wedge triangle(x_1)\wedge square(x_2)\ \wedge circle(x_3) \quad (27')$$

$$(\exists x,y)L(Tom,x,y,Tom,A_{12},G_s,\_)\Pi L(Tom,x,In,In,A_{44},G_t,K_{9IM})\ \wedge ISR(x)\ \wedge room(y) \quad (28)$$

$$(\exists x,y)L(Tom,x,Tom,y,A_{12},G_s,\_)\Pi L(Tom,x,Cont,Cont,A_{44},G_t,K_{9IM})\wedge ISR(x)\ \wedge room(y) \quad (28')$$

$$(\exists x,y,p,q)L(Tom,Tom,p,q,A_{12},G_t,\_)\Pi L(Tom,x,y,Tom,A_{12},G_s,\_)$$
$$\Pi L(Tom,x,In,Dis,A_{44},Gt,K_{9IM})\ \wedge ISR(x)\ \wedge room(y)\ \wedge p\neq q \quad (29)$$

The rigid topology between two objects as in the 9-intersection model must be determined with the perfect knowledge of their insides, outsides and boundaries. Ordinary people, however, can comment on matters without knowing all about them. This is the very case when they encounter an unknown object too large to observe at a glance just like a road in a strange country. For example, Fig.8-a shows such a path viewed from the sky that is partly hidden by the woods. In this case, the topological relation between the path as a whole and the swamp/woods depends on how the path starts and ends in the woods, but people could utter such sentences as S10 and S11 about this scene. Actually, these sentences refer to such spatial events that reflect certain temporal changes in the topological relation between the swamp/woods and FAO running along the path. Therefore, their conceptual descriptions are to be given as (30) and (31), respectively. For another example, Fig.8-b shows a more complicated spatial event in topology that can be formulated as (32) and could be verbalized as S12.
(S10) The path enters the swamp/woods.
(S11) The path exits the swamp/woods.

$$(\exists x,y,z)L(\_,z,p,q,A_{12},G_s,\_)\Pi L(\_,x,y,z,A_{12},G_s,\_)\Pi$$
$$L(\_,x,Dis,In,A_{44},G_s,K_{9IM})\ \wedge ISR(x)\ \wedge\{swamp(y)/woods(y)\}\ \wedge path(z)\ \wedge p\neq q \quad (30)$$

$$(\exists x,y,z)L(\_,z,p,q,A_{12},G_s,\_)\Pi L(\_,x,y,z,A_{12},G_s,\_)\Pi$$
$$L(\_,x,In,Dis,A_{44},G_s,K_{9IM})\ \wedge ISR(x)\ \wedge\{swamp(y)/woods(y)\}\ \wedge path(z)\ \wedge p\neq q \quad (31)$$

(S12) The path cuts the swamp twice (as shown in Fig.8-b), passing p_1 outside, p_2 inside, p_3 outside, p_4 inside and p_5 outside the swamp on the way.

$$(\exists x,y,z,p_1,\ldots,p_5)L(\_,z,y,x,A_{12},G_s,\_)\Pi((L(\_,x,p_1,p_2,A_{12},G_s,\_)$$
$$\Pi(L(\_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(\_,x,p_2,p_3,A_{12},G_s,\_)\Pi L(\_,z,In,Dis,A_{44},G_s,K_{9IM}))$$
$$\bullet(L(\_,x,p_3,p_4,A_{12},G_s,\_)\Pi L(\_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(\_,x,p_4,p_5,A_{12},G_s,\_)\Pi$$
$$L(\_,z,In,Dis,A_{44},G_s,K_{9IM})))\wedge path(x)\wedge swamp(y)\wedge ISR(z) \quad (32)$$

Lastly, consider such somewhat complicated sentences as S13 and S14. The underlined parts are deemed to refer to some events neglected in time and in space, respectively. These events correspond with skipping of FAOs and are called 'Temporal Empty Event' and 'Spatial Empty Event', denoted by 'ε_t' and 'ε_s' as EEs with g=G_t and g=G_s at (11), respectively. The concepts of S13 and S14 are given by (33) and (34), where 'A_{15}' and 'A_{17}' represent the attribute 'Trajectory' and 'Mileage', respectively.

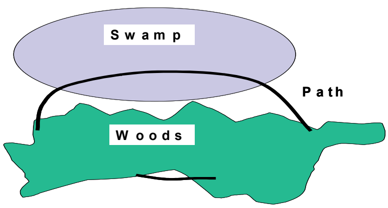
(S13) The *bus* runs 10km straight east from A to B, and *after a while*, at C it meets the street with the sidewalk.

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

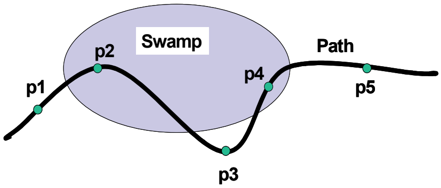
$$(\exists x,y,z,p,q)(L(\_,x,A,B,A_{12},G_t,\_)\Pi L(\_,x,0,10km,A_{17},G_t,\_)\Pi$$
$$L(\_,x,Point,Line,A_{15},G_t,\_)\Pi L(\_,x,East,East,A_{13},G_t,\_)) \bullet \varepsilon_t \bullet \quad (33)$$
$$(L(\_,x,p,C,A_{12},G_t,\_)\Pi L(\_,y,q,C,A_{12},G_s,\_)\Pi L(\_,z,y,y,A_{12},G_s,\_))$$
$$\wedge bus(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q$$

$$(\exists x,y,z,p,q)(L(\_,x,A,B,A_{12},G_s,\_)\Pi L(\_,x,0,10km,A_{17},G_s,\_)\Pi$$
$$L(\_,x,Point,Line,A_{15},G_s,\_)\Pi L(\_,x,East,East,A_{13},G_s,\_)) \bullet \varepsilon_s \bullet \quad (34)$$
$$(L(\_,x,p,C,A_{12},G_s,\_)\Pi L(\_,y,q,C,A_{12},G_s,\_)\Pi L(\_,z,y,y,A_{12},G_s,\_))$$
$$\wedge road(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q$$

From the viewpoint of cross-media reference as integrated multimedia understanding, the formula (34) can refer to such a spatial event depicted as the still picture in Fig.9 while (33) can be interpreted into a motion picture.



(a)



(b)

Fig. 8. Delicate topological relations: (a) the path partially hidden by the woods and (b) the path winding inside-outside-inside-outside of the swamp

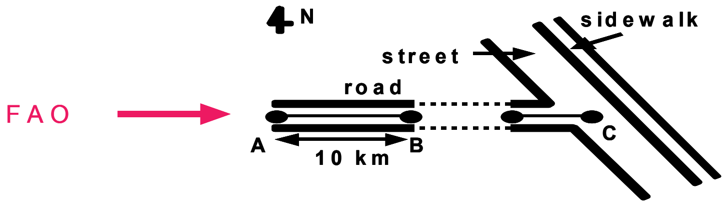


Fig. 9. Pictorial interpretation of the formula (34) and the movement of FAO involved

3.2 Conceptualization

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 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 10 illustrates the mental process of conceptualization on the word 'conveyance', where a set of its referents ***Sr*** (={*Matter_1*,…, *Matter_n*}) are generalized by abstraction and formulated as (35). The underlined part of this formula implies that matter 'z' includes two matters 'x' and 'y' in its 'Place of existence (A_{01})'. 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 (A_{01})' and 'Physical location (A_{12})' 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'.

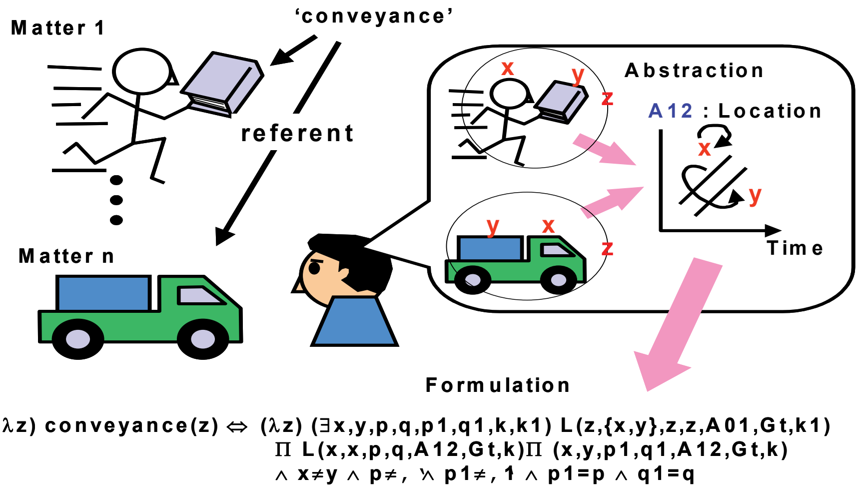


Fig.10. Conceptualization: the process of abstraction on referents of the word 'conveyance'.

$$(\lambda x)\text{conveyance}(z) \Leftrightarrow (\lambda z)(\exists x,y,p,q,p_1,q_1,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}$$

$$\Pi L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p_1,q_1,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q\wedge p_1\neq q_1\wedge p_1=p\wedge q_1=q$$

$$\Leftrightarrow (\lambda z)(\exists x,y,p,q,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}$$

$$\Pi L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q$$

$$\Leftrightarrow (\lambda z)(\exists x,y,p,q,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}\Pi L(x,\{x,y\},p,q,A_{12},G_t,k) \wedge x\neq y \wedge p\neq q \tag{35}$$

For another example, the matter called 'snow' can be conceptualized as (36), where '_', as defined by (37), 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 2 for undefined attributes.)

$$(\lambda x)\text{snow}(x)\Leftrightarrow(\lambda x)(\exists x_1,x_2,\ldots)((L(\_,x,x_1,x_1,A_{41},G_t,\_)\Pi L(\text{Earth},x,\text{Sky},\text{Earth},A_{12},G_t,\_))$$

$$\wedge L(\_,x,x_1,x_2,A_{41},G_t,\_) \wedge\text{powder}(x_1)\wedge \text{ice}(x_1)\wedge\text{water}(x_2) \wedge\ldots) \tag{36}$$

$$L(\ldots,\_,\ldots) \leftrightarrow (\exists x)L(\ldots,x,\ldots) \tag{37}$$

For a more complicated example, the concept of 'umbrella' can be represented as (38), reading 'At raining, a human puts an umbrella in line between rain and himself/herself in order not to get wetter, …'

By the way, the concepts of 'rain' and 'wind' can be given as (39) and (40), 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.

$$(\lambda x)\text{umbrella}(x) \Leftrightarrow (\lambda x)(\exists x,x_1,x_2,y,p,q_1,q_2,\ldots)(L(\_,x_1,p,p,A_{13},G_t,\_)\Pi$$

$$((L(\_,y,x_1,x,A_{12},G_s,\_)\bullet L(\_,y,x,x_2,A_{12},G_s,\_))\Pi\ L(\_,y,p,p,A_{13},G_s,\_)\Pi$$

$$L(x_1,x_2,q_1,q_2,A_{25},G_t,\_)\wedge ISR(y)\wedge rain(x_1)\wedge human(x_2)\wedge\sim(q_1<q_2)\ldots) \tag{38}$$

$$(\lambda x)\text{rain}(x) \Leftrightarrow (\lambda x)(\exists x_1,x_2,x_3,x_4,p,q,\ldots)L(\_,x,x_1,x_1,A_{41},G_t,\_)\Pi$$

$$L(Earth,x,Sky,Earth,A_{12},G_t,\_)\Pi L(x,x_2,p,q,A_{25},G_t,\_)\Pi L(x_3,x_4,x,x,A_{19},G_t,x_3)$$

$$\wedge water(x_1)\wedge object(x_2)\wedge human(x_3)\wedge umbrella(x_4)\wedge(p<q)\ldots \tag{39}$$

$$(\lambda x)\text{wind}(x) \Leftrightarrow (\lambda x)(\exists x_1,x_2,p,q,\ldots)L(\_,x,x_1,x_1,A_{41},G_t,\_)\wedge air(x_1)\wedge$$

$$(L(x,x_2,p,q,A_{13},G_t,\_)\wedge rain(x_2)\ldots \tag{40}$$

3.3 Knowledge of word meanings

A word meaning M_W is defined as a pair of 'Concept Part (C_p)' and 'Unification Part (U_p)' and is formulated as (41).

$$M_W = [C_p{:}U_p] \tag{41}$$

The C_p of a word W is an $\boldsymbol{L_{md}}$ expression as its concept 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 (42).

$$[(\lambda x,y)(\exists p,q,k)L(x,\{x,y\},p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q{:}\ ARG(Dep.1,x);\ ARG(Dep.2,y);] \tag{42}$$

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 (43).

$$(\exists y,p,q,k)\ L(Mary,\{Mary,y\},p,q,A_{12},G_t,k)\wedge Mary\neq y\wedge p\neq q\wedge book(y) \tag{43}$$

Figure 11 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 (44).

$$[(\lambda x,y)(\exists p_1,z,p_3,g,k,p_4,k_0)\ (\underline{L(x,y,p_1,z,A_{12},g,k)}\bullet L(x,y,z,p_3,A_{12},g,k))\Pi$$

$$L(x,y,p_4,p_4,A_{13},g,k_0)\wedge p_1\neq z\wedge z\neq p_3{:}\ ARG(Dep.1,z);$$