Logical Formalization of Human Intuitive Mental Imagery for Spatiotemporal Language Understanding

ROJANEE KHUMMONGKOL*, MASAO YOKOTA**
Department of Intelligent Information System Engineering, Graduate School
Fukuoka Institute of Technology
3-30-1 Wajirohigashi, Higashi-ku, Fukuoka 811-0295
JAPAN
*bd14002@bene.fit.ac.jp, **yokota@fit.ac.jp

Abstract: - People would employ intuitive spatial relations such as ‘at’, ‘around’, ‘along’, etc. rather than the rigid global coordinates provided by GPS when they refer to physical locations in their casual communication. These word concepts, however, are considered to reflect human cognitive propensities toward the external world. Therefore they can be subjective to each human individual but as well can be common to the human beings in the macroscopic qualities of their conceptualization. This paper describes a method for systematic representation and computation of human intuitive spatiotemporal knowledge based on a mental image model and its application to natural language understanding.

Key-Words: - Natural language understanding, Spatiotemporal knowledge, Knowledge representation, Mental image

1 Introduction

Nowadays, location measurement systems like GPS are widely utilized to acquire locations in the global coordinates system and have already been applied to location-aware systems of practical use such as vehicle-navigation systems. It is, however, not very common for people to refer to these coordinates but intuitive spatial relations such as ‘at’, ‘around’, ‘along’, etc. in their casual communication. Such natural concepts, however, are considered to reflect human intuitive mental activities toward the world. Therefore they can be subjective to each human individual but as well can be common to the human beings in the macroscopic qualities of their conceptualization. This paper describes a method for systematic representation and computation of human intuitive spatiotemporal (i.e., 4D) concepts as mental imagery and its application to natural language understanding (NLU).

As well known, people do not perceive the external world so as it is, which naturally leads to human-specific cognition and conception of the external world. For example, as shown in Fig.1a and 1b, people often perceive continuous forms among separately located objects so called spatial gestalts in the field of psychology and refer to them by such expressions as S1 and S2, respectively.

(S1) Five disks are in *line*.

(S2) Nine disks are placed in *the shape of X*.

(a) (b)

Fig. 1. Gestalts perceived among multiple disks.

For another example, people would intuitively and easily understand the following expressions S3 and S4 so that they should describe the same scene in the external world. This is also the case for S5 and S6.

(S3) The path *sinks* to the stream.
(S4) The path *rises* from the stream.
(S5) Route A and Route B *meet* at the city.
(S6) Route A and Route B *separate* at the city.

Apparently, these expressions do not reflect much the purely objective geometrical relations, exclusively treated in conventional artificial intelligence (AI) based on numerical measures [1,2,3,4], but very much human mental activity at cognition of the objects involved and it seems extremely difficult for robots to reach such a paradoxical understanding in a systematic way. From the semantic viewpoint, as easily imagined by these figures, spatial expressions have the virtue of relating in some way to visual scenes being described. Therefore, their semantic descriptions can be grounded in perceptual representation as mental image, possibly, cognitively inspired and coping with all kinds of spatial expressions including such verb-centered ones as S3-S6 as well as preposition-centered ones.

Mental Image Directed Semantic Theory (MIDST) [5,6] has already proposed a model of human active (or attention-guided) 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 Change Events' and 'Spatial Change Events'. These are defined as temporal and spatial changes (or constancies) in certain attributes of physical objects, respectively, with S3-S6 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} (Language for Mental-image Description). This language is employed for many-sorted predicate logic and can provide 4D expressions with computable semantic descriptions as their perceptual representations.

The remainder of this paper is organized as follows. Section 2 presents a formal system for representation and computation of human mental imagery in L_{md}. Section 3 describes how to formulate human knowledge as mental imagery of space and time in L_{md} and section 4 applies it to a conversation management system under development. Finally, section 5 discusses and concludes this paper.

2 Formal System

A formal system is defined as a pair of a formal language and a deductive system consisting of the axioms and inference rules employed for theorem derivation. L_{md} is a formal language for many-sorted predicate logic with 5 types of terms specific to the mental image model. Therefore, the deductive system intended here is to be based on the deductive apparatus for predicate logic.

The symbols of L_{md} for the deductive system are listed as (i)-(xi) below. These symbols are possibly subscripted just like A_{01}, G_s, etc.

(i) logical connectives : $\sim, \wedge, \vee, \supset, \equiv$
(ii) quantifiers : $\forall, \exists$
(iii) auxiliary constants : ., (,)
(iv) sentence variables : χ
(v) individual variables
  a. matter variables : x, y, z
  b. attribute variables : a
  c. value variables : p, q, r, s, t
  d. pattern variables : g
  e. standard variables : k
(vi) sentence constants : N
(vii) predicate constants : L, =, $\neq$, >, < , and others to be introduced where needed
(viii) individual constants
  a. matter constants : to be introduced where needed
  b. attribute constants : A, B
  c. value constants : to be introduced where needed
  d. pattern constants : G
  e. standard constants : K
(ix) function constants: arithmetic operators such as +, -, etc. and others to be introduced where needed
(x) meta-symbols: $\Leftrightarrow$, $\rightarrow$, $\leftrightarrow$, and others to be introduced where needed
(xi) others: to be defined by the symbols above

The system is a many-sorted predicate logic with five kinds of individuals employed for one special predicate constant '*L*' so called 'Atomic Locus'. Except this point, the syntactic rules and the theses of the system are the same as those of the conventional predicate logic. The predicate '*L*' is such a seven-place predicate as is given by expression (1).

$$L(\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7) \quad (1)$$

The formula (1) is a well-formed formula (i.e. wff) called 'Atomic locus formula' if and only if the conditions (a)-(g) below are satisfied. A well-formed formula consisting of atomic formulas and logical connectives is called simply 'Locus formula'.

a. ω_1 is a matter term (variable or constant)
b. ω_2 is a matter term
c. ω_3 is a value or a matter term
d. ω_4 is a value or a matter term
e. ω_5 is an attribute term
f. ω_6 is a pattern term
g. ω_7 is a standard (or matter) term

The intuitive interpretation of (1) is given as follows:

"Matter ω_1 causes Attribute ω_5 of Matter ω_2 to keep ($\omega_3 = \omega_4$) or change ($\omega_3 \neq \omega_4$) its Values temporally ($\omega_6 = G_t$) or spatially ($\omega_6 = G_s$) over a certain absolute time-interval, where Values ω_3 and ω_4 are relative to Standard ω_7."

It is notable that Matter terms placed at ω_3, ω_4 or ω_7 represent their values in each place at the time or over the time-interval. Furthermore, when it is not significant to discern variables at ω_1, ω_3, ω_4 or ω_7, anonymous variables, usually symbolized as '_', are employed in their places (See (5) for example).

In the interpretation above, when $\omega_6=G_t$, the locus indicates monotonic change (or constancy) of the attribute in time domain, and when $\omega_6=G_s$, that in space domain. The former is called 'temporal change event' and the latter, 'spatial change event'. For example, the motion of the 'bus' represented by S7 is a temporal change event and the ranging or extension of the 'road' by S8 is a spatial change event whose meanings or concepts are formulated as (2) and (3), respectively, where 'A_{12}' denotes the attribute 'Physical Location'. These two formulas are different only at the term 'Pattern' (=ω_6).

(S7) The bus runs from Tokyo to Osaka.
(S8) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,\mathrm{Tokyo},\mathrm{Osaka},A_{12},G_t,k)\wedge \mathrm{bus}(y) \quad (2)$$

$$(\exists x,y,k)L(x,y,\mathrm{Tokyo},\mathrm{Osaka},A_{12},G_s,k)\wedge \mathrm{road}(y) \quad (3)$$

The deductive system employs 'tempo-logical connectives (TLCs)' with which to represent both temporal and logical relations between two loci over certain time-intervals. The definition of a tempo-logical connective C_i is given by (4), where τ_i, χ and C refer to one of ***pure*** temporal relations indexed by an integer '*i*', a locus, and an ordinary binary logical connective such as the conjunction '$\wedge$', respectively. The definition of each τ_i discriminates 13 types of temporal relations by the integer suffix '*i*' ranging from –6 to 6 [7]. 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.

$$\chi_1 \, C_i \, \chi_2 \Leftrightarrow (\chi_1 \, C \, \chi_2) \wedge \tau_i(\chi_1, \chi_2), \quad (4)$$

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

3 Formulation of Subjective 4D Knowledge

MIDST hypothesizes that the difference between temporal and spatial change event concepts can be attributed to the relationship between the Attribute Carrier (AC) (i.e. the matter at ω_2) 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 change event but *runs* about on the AC in a spatial change event. Consequently, the *bus* and FAO move together in the case of S7 while FAO solely moves along the *road* in the case of S8. That is, ***all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal change events of FAO***.

Therefore, an event expressed in $\boldsymbol{L_{md}}$ 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 the FAO as illustrated in Fig.2a and whose conceptual descriptions are given as (5) and (6), respectively, where 'A_{13}', '↑' and '↓' refer to the attribute 'Direction' and its values 'upward' and 'downward', respectively. Such a fact is generalized as '***Postulate of Reversibility of Spatial Change Event (PRS)***'. This postulate is also valid for such a pair of S5 and S6 as interpreted approximately into (7) and (8), respectively. These pairs of conceptual descriptions are called equivalent in the PRS, and the paired sentences are treated as paraphrases each other.

$$(\exists y,p,z)L(\_,y,p,z,A_{12},G_s,\_)\Pi L(\_,y,\downarrow,\downarrow,A_{13},G_s,\_) \wedge \mathrm{path}(y)\wedge \mathrm{stream}(z)\wedge p\neq z \quad (5)$$

$$(\exists y,p,z)L(\_,y,z,p,A12,G_s,\_)\Pi L(\_,y,\uparrow,\uparrow,A_{13},G_s,\_) \wedge \mathrm{path}(y)\wedge \mathrm{stream}(z)\wedge p\neq z \quad (6)$$

$$(\exists p,y,q)L(\_,\mathrm{Route\_A},p,y,A_{12},G_s,\_) \Pi L(\_,\mathrm{Route\_B},q,y,A_{12},G_s,\_)\wedge \mathrm{city}(y)\wedge p\neq q \quad (7)$$

$$(\exists p,y,q)L(\_,\mathrm{Route\_A},y,p,A_{12},G_s,\_) \Pi L(\_,\mathrm{Route\_B},y,q,A_{12},G_s,\_)\wedge \mathrm{city}(y)\wedge p\neq q \quad (8)$$

For another example of spatial change event, Fig. 2b concerns the perception of the formation of multiple objects, where FAO runs along an imaginary object so called '***Imaginary Space Region (ISR)***'. This event can be verbalized as S9 using the preposition 'between' and formulated as (9) or (9'), corresponding also to such concepts as 'row', 'line-up', etc. Employing ISRs and the 9-intersection model [8], topological relations between two objects can be formulated in such expressions as (10) or (10') for S10, and (11) for S11, 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 transposes each other.

(S9) The square is between the triangle and the circle.

$$(\exists x_1,x_2,x_3,y,p,q)(L(\_,y,x_1,x_2,A_{12},G_s,\_)\ \Pi L(\_,y,p,p,A_{13},G_s,\_))\bullet(L(\_,y,x_2,x_3,A_{12},G_s,\_)\ \Pi L(\_,y,q,q,A_{13},G_s,\_))\wedge ISR(y)\wedge p=q\wedge triangle(x_1)\wedge square(x_2)\wedge circle(x_3) \qquad (9)$$

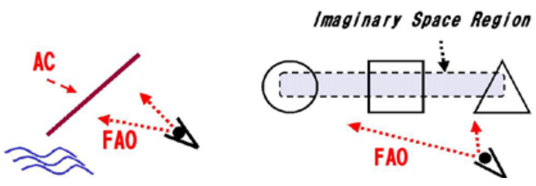

Fig. 2. FAO movements:
'slope' (a) and 'row' (b) as spatial events.

$$(\exists x_1,x_2,x_3,y,p)(L(\_,y,x_1,x_2,A_{12},G_s,\_)\bullet L(\_,y,x_2,x_3,A_{12},G_s,\_))\Pi L(\_,y,p,p,A_{13},G_s,\_)\wedge ISR(y)\wedge triangle(x_1)\wedge square(x_2)\wedge circle(x_3) \qquad (9')$$

(S10) Tom is in the room.

$$(\exists x,y)L(Tom,x,y,Tom,A_{12},G_s,\_)\ \Pi L(Tom,x,In,In,A_{44},G_t,K_{9IM})\wedge ISR(x)\wedge room(y) \qquad (10)$$

$$(\exists x,y)L(Tom,x,Tom,y,A_{12},G_s,\_)\ \Pi L(Tom,x,Cont,Cont,A_{44},G_t,K_{9IM})\wedge ISR(x)\wedge room(y) \qquad (10')$$

(S11) Tom exits the room.

$$(\exists x,y,p,q)L(Tom,Tom,p,q,A_{12},G_t,\_)\ \Pi L(Tom,x,y,Tom,A_{12},G_s,\_)\ \Pi L(Tom,x,In,Dis,A_{44},Gt,K_{9IM})\wedge ISR(x)\wedge room(y)\wedge p\neq q \qquad (11)$$

The mathematically rigid topology between two objects must be determined with the perfect knowledge of their insides, outsides and boundaries [8]. Ordinary people, however, would often 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. 3a 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 S12 and S13 about this scene. Actually, these sentences refer to such events that reflect certain temporal changes in the topological relation between the swamp/woods and the FAO running along the path.

Therefore, their conceptual descriptions are to be given as (12) and (13), respectively. For another example, Fig. 3b shows a more complicated spatial event in topology that can be formulated as (14) and can be verbalized as S14.

(S12) The path goes into the swamp/woods.
(S13) The path comes out of the swamp/woods.

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

$$(\exists x,y,z)L(\_,z,p,q,A_{12},G_s,\_)\ \Pi L(\_,x,y,z,A_{12},G_s,\_)\ \Pi L(\_,x,In,Dis,A_{44},G_s,K_{9IM})\wedge ISR(x)\wedge\{swamp(y)/woods(y)\}\wedge path(z)\wedge p\neq q \qquad (13)$$

(S14) The path cuts the swamp twice as shown in Fig. 3b, passing p_1 outside, p_2 inside, p_3 outside, p_4 inside and p_5 outside the swamp on the way.

$$(\exists x,y,z,p_1,\ldots,p_5)L(\_,z,y,x,A_{12},G_s,\_)\ \Pi((L(\_,x,p_1,p_2,A_{12},G_s,\_)\ \Pi(L(\_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(\_,x,p_2,p_3,A_{12},G_s,\_)\ \Pi L(\_,z,In,Dis,A_{44},G_s,K_{9IM}))\bullet(L(\_,x,p_3,p_4,A_{12},G_s,\_)\ \Pi L(\_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(\_,x,p_4,p_5,A_{12},G_s,\_)\ \Pi L(\_,z,In,Dis,A_{44},G_s,K_{9IM})))\wedge path(x)\wedge swamp(y)\wedge ISR(z) \qquad (14)$$

Conventional approaches to spatial (i.e., 3D) language understanding have inevitably employed a tremendously great number of axioms such as (15). It is noticeable that these axioms are part of the definition of 'between' valid only for verbalized directions such as 'left' and 'above' and that actually much more axioms should be needed for other directions such as 'before' and 'behind'.

$$(\forall x,y)right(x,y)\equiv left(y,x)$$
$$(\forall x,y)above(x,y)\equiv under(y,x)$$

$(\forall x,y,z)\mathrm{above}(y,x)\&\mathrm{above}(x,z)\supset\mathrm{between}(x,y,z)$
$(\forall x,y,z)\mathrm{right}(y,x)\&\mathrm{right}(x,z)\supset\mathrm{between}(x,y,z)$ (15)

Distinguishably, MIDST gives the definition of 'between' in such a simple and language-free formula as the underlined part of (9') and moreover that is applicable to every direction, whether or not verbalized. The concepts of 40 English prepositions, so-called, spatial prepositions such as 'along' were analyzed and formulated in accordance with MIDST. To be most remarkable, the concepts of spatial prepositions are defined as 4D images in MIDST but not as 3D (=4D minus 'time') images in conventional approaches.

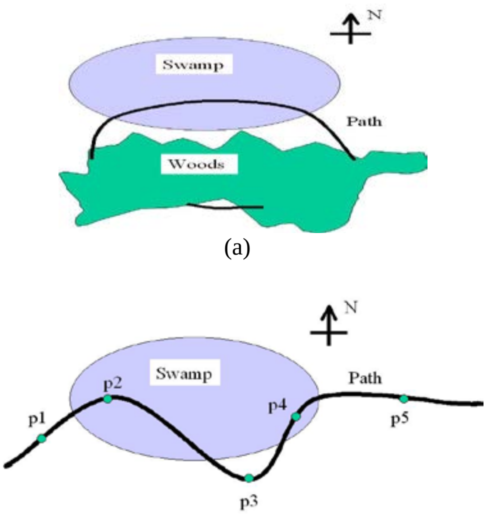

Fig. 3. Delicate topological relations: (a) path partially hidden by woods and (b) path winding in-out-in-out of swamp.

4 Application to Natural Language Understanding

The authors have been developing a conversation management system (CMS) in Python as shown in Fig. 4 for simulating human-robot interaction in NL [9]. The implemented AI named *Anna* understands NL utterances in text by a person and each person as her dialogue partner is to be played as by the system user named *Taro.* Anna is assumed as a lady-shaped helper robot for Taro, a physically handicapped elderly man.

Consider such a question as S14 to a certain NLU system from its human user. Then, this question can be read roughly as (16), where ➔ means that the right hand is deduced from the left hand by applying some pieces of knowledge.

(S14) When Tom drives with Mary, does she move?

?drive(Tom)&with(Tom,Mary)
➔ move(Mary) (16)

For conventional NLU systems to answer such a question correctly, some special postulate like (17) should be needed, where → means that if the left hand is true then the right hand is true, too.

drive(x)&with(x,y) → move(y) (17)

However, people can easily answer 'yes' without (17). How do they do that? They must employ their mental image evoked by their own experiences. Anna can imitate such a human thinking process. Her understanding of S16 can be depicted as Fig. 5, where Loc = (A_{12},G_t,k) and the $\boldsymbol{L_{md}}$ expression reads 'Tom stays himself in the car, and simultaneously (Π), Tom moves the car form P to Q, and simultaneously, Tom stays Mary at his place', and therefore (➔) 'Tom moves himself from P to Q, and simultaneously, Tom stays Mary at his place', and therefore, 'Tom moves himself from P to Q, and simultaneously, Tom moves Mary from P to Q', and therefore, 'Tom moves Mary from P to Q', and therefore 'Mary moves from P to Q, (consequently)'.

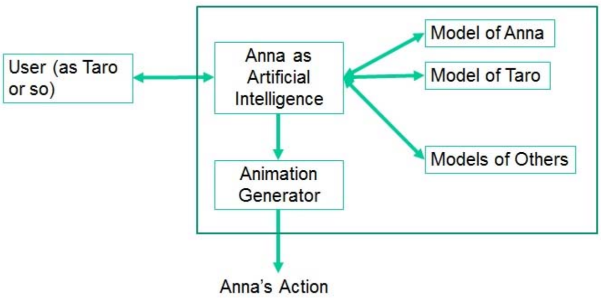

Fig. 4. Conversation management system.

Fig. 5. Anna's thinking process for S14.

5 Discussion and Conclusion

Awareness computing [10], viewed from robotic interaction with the physical world, is conventionally conceived as sensory data computing driven by certain heuristics cognitively motivated and intended to reduce their computational cost [11]. Its current goal is to build systems that are not necessarily required to comprehend sensory data totally but to solve many practical problems. This paper considered awareness computing based on a formal language $\boldsymbol{L_{md}}$ designed for total comprehension of sensory data, focusing on location-aware computing based on human subjective 4D knowledge as mental imagery in $\boldsymbol{L_{md}}$.

We have analysed a considerable number of spatial 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, and found that almost all the concepts of spatial events can be defined in exclusive use of 5 kinds of attributes for FAOs, namely, Physical location (A_{12}), Direction (A_{13}), Trajectory (A_{15}), Mileage (A_{17}) and Topology (A_{44}). This implies that location-aware systems based on the formal system are very feasible in the respect of the size of knowledge to be installed. The inferences in $\boldsymbol{L_{md}}$ are based on simple and general rules about atomic loci and therefore CMS works feasibly in Python except for computational cost in the animation generator. Fig. 6 and 7 in Appendix show examples of session between Anna (=CMS) and Taro (=User).

Recently, deep neural networks (DNNs) have been achieving remarkable performance on various pattern-recognition tasks, especially visual classification problems. Simultaneously, interesting differences have been pointed out between human vision and current DNNs arousing questions about the generality of DNN computer vision because there exist images that are completely unrecognizable to humans, but that DNNs believe to be recognizable objects with 99.99% confidence [12]. However, DNNs as is cannot accept any feedback from higher level cognition such as reasoning in order to calibrate misclassification because of lacking any immediate means to convert knowledge representation (as awareness of misclassification by reasoning) into weight sets for connectionism. This may lead to such a claim that artificial intelligence (AI) should be more cognitive [13,14]. We believe that this research as well can give a good suggestion about how machine learning (ML) should acquire knowledge to be available for higher level cognition such as abstract reasoning. Our future work will include systematic incorporation of ML into our conversation management system for automatic acquisition of natural concept and knowledge.

References:

[1] G.D. Logan and D.D. Sadler, *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, 1996.

[2] A. Carstensen, E. Kon, and T. Regier, *Testing a rational account of pragmatic reasoning: The case of spatial language*, Proceedings of the 36 th Annual Meeting of the Cognitive Science Society. Austin, TX, 2009–2013, 2014.

[3] D. Golland, P. Liang, and D. Klein, *A Game-Theoretic Approach to Generating Spatial Descriptions*, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 410-419, Cambridge, MA, October 2010.

[4] T. D. Ullman, Y. Xu, and N. D. Goodman, *The Pragmatics of Spatial Language*, Proceedings of the 38th Annual Conference of the Cognitive Science Society, 2016.

[5] M. Yokota, *An Approach to Natural Language Understanding Based on Mental Image Model*, In Sharp,B. (ed.), Natural Language Understanding and Cognitive Science, pp.22-31, INSTICC PRESS, 2005.

[6] M. Yokota, *Aware Computing in Spatial Language Understanding Guided by Cognitively Inspired Knowledge Representation*, Applied Computational Intelligence and Soft Computing,, vol. 2012 Article ID 184103, 10 pages, doi:10.1155/2012/184103, 2012.

[7] J.F. Allen, *Towards a general theory of action and time*, Artificial Intelligence, 23-2, pp.123-154, 1984.

[8] M. Egenhofer, *Point-set topological spatial relations*, Geographical Information Systems, 5-2, pp.161-174, 1991.

[9] R. Khummongkol and M. Yokota, *Computer Simulation of Human-Robot Interaction in Natural Language*, ISAROB'16, 2016.

[10] Q. Zhao (ed.), *Proceedings of the IEEE 7th International Conference of Awareness Science and Technology*, Qinhuangdao, China, 2015.

[11] R. A. Brooks, *A robust layered control system for a mobile robot*, IEEE Journal of Robotics and Automation, Vol.RA-2, 14-23, 1986.

[12] A. Nguyen, J. Yosinski, and J. Clune, *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, CVPR'15, 2015.

[13] Z. Kishimoto, *What's Wrong Current Approaches to AI?*, http://tek-

tips.nethawk.net/what%E2%80%99s-wrong-current-approaches-to-ai/ (Accessed 2016-09-14)

[14] B. M. Lake et al., *Building Machines That Learn and Think Like People*, CBMM Memo 46, 55 pages, 2016.

Appendix:

```
C:\Python27\python.exe C:/Users/20150714yo/PycharmProjects/iWork/3DProject_new/main.py
************* Welcome to Anna's world **************
*** To exit the program, please press 'q' or 'Q' ***


Your command: When Tom drives a car to the flower_bed with Mary, does she move?

    === Output(s) from the processes ===
    ---Final locus---
    final locus = [[['sand'], ['sand'], ['tom', 'tom', 'car', 'car', 'a', 'v', '+', 0], ['tom
', 'car', 'p', 'flower bed', 'a', 'v', '+', 0], ['tom', 'mary', 'tom', 'tom', 'a', 'pp', '+', 0
]], [['x', 'she', 'p', 'q', 'a', 'v', '+', 1]]]

    ---Final locus after applied postulate and inference rules---
    final locus = [[['sand'], ['sand'], ['sand'], ['sand'], ['sand'], ['tom', 'tom', 'car', 'car
', 'a', 'v', '+', 0], ['tom', 'car', 'p', 'flower_bed', 'a', 'v', '+', 0], ['tom', 'mary', 'tom
', 'tom', 'a', 'pp', '+', 0], ['tom', 'tom', 'p', 'flower bed', 'a', 'v', '+', 0], ['tom', 'mary
', 'car', 'car', 'a', 'pp', '+', 0], ['tom', 'mary', 'p', 'flower_bed', 'a', 'pp', '+', 0]], [['
x', 'mary', 'p', 'q', 'a', 'v', '+', 1]]]

Anna: Yes, mary does.
Anna: From apartment,
     after that, move to the house,
     after that, move to the yard,
     after that, move to the bridge,
     then, go to the fountain,
     after that, move to the table,
     then, go to the stair,
     finally mary will arrive at the flower_bed.
```

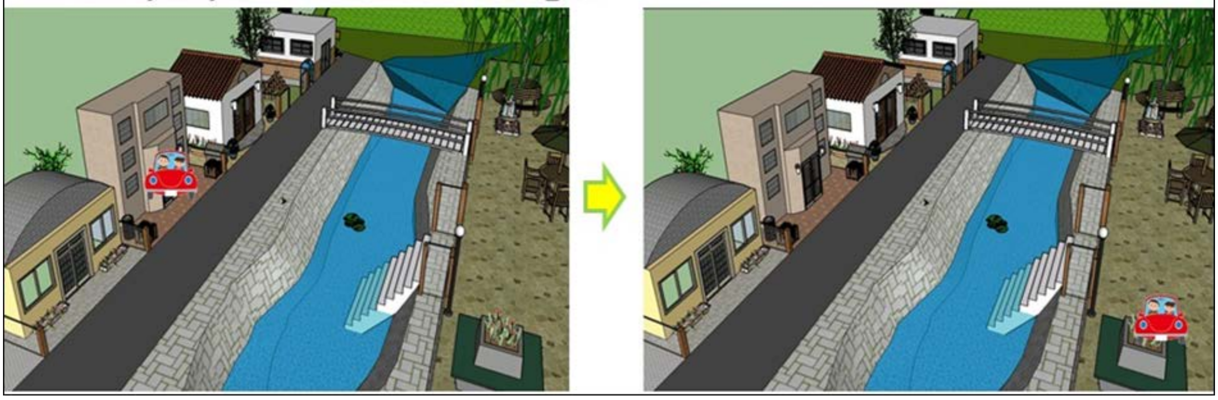

Fig. 6. Example of conversation between Anna and Taro and snapshots of generated animation:
The generated animation shows the scene of Tom's driving from home to the flower-bed with Mary.

```
Your command: Please take me to the fountain.

Anna: Now! Let's go to fountain

Your command: How can I go to the apartment?

Anna: From fountain.
Anna: after that, walk to the bridge.
Anna: after that, walk to the yard.
Anna: then, go to the house.
Anna: then, go to the apartment.

Your command: Where is the post_office?

Anna: The post_office is to the right of the yard.

Your command: The road runs along the river.

Anna: I see, and then??

Your command: How does the river run?

Anna: The river runs from north to south.

Your command: q

Process finished with exit code 0
```

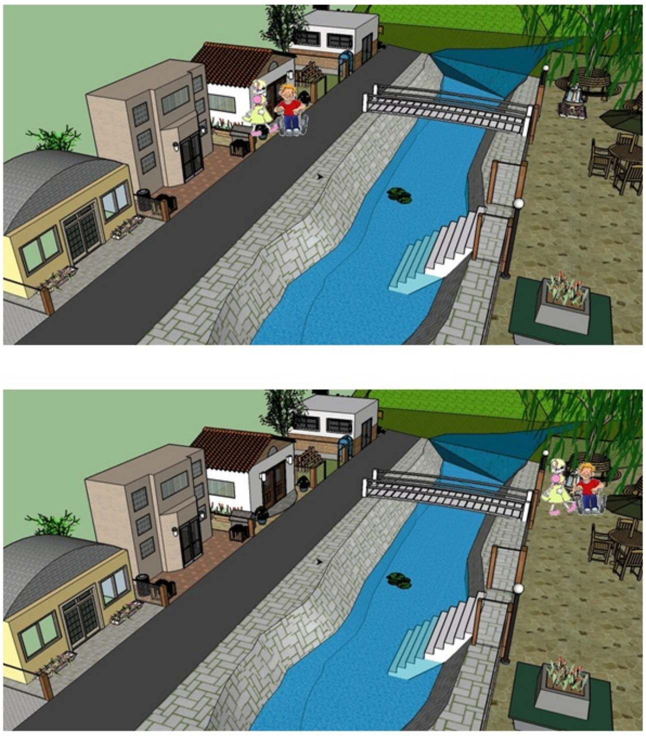

Fig. 7. Another example of conversation between Anna and Taro and snapshots of generated animation: The generated animation shows the scene of Anna's taking Taro from home to the fountain.