We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Connectives Acquisition in a Humanoid Robot Based on an Inductive Learning Language Acquisition Model

Dai Hasegawa, Rafal Rzepka & Kenji Araki
Hokkaido University
Japan

1. Introduction

1.1 Background

Humanoid robots, like the Honda's ASIMO[1], have been developed and such machines will be expected to help people to do a variety of tasks in every day life. Then robots that exist in the same dynamic environment as humans should be able to interact with humans using natural language.

However, when it comes to developing robots that understand language and work in dynamic environment, there are two problems to be solved in usual methods of the natural language processing.

One problem is that of the language grounding which was also pointed out by Roy (Roy, 2003). In natural language processing systems, meanings of words are defined by other words circularly. That is, the words are not connected to objects, movements, colors or other physical features in the real world. However, connecting words to the real world is needed when a robot is to perform a concrete action following a human's utterance. For example, when a person orders a robot to "kick the red ball", if the word "red" is not connected to a specific color representation in accordance with the real world then the robot cannot realize the order.

Another problem is that the top-down approach, which is very common in natural language processing systems, cannot deal with a dynamic environment. For example, SHRDLU (Winograd, 1972) is a system that connects words to a virtual world where some blocks exist. This system can understand utterances that refer to the virtual world such as "put the red cylinder on the green box" and can perform such actions when told, because the designer describes knowledge about the limited and static virtual world to the system in advance. On the other hand, SHRDLU cannot reply to utterances that refer to things out side of the virtual world such as "is the weather fine today?". Hence, the top-down approach is not appropriate to design a robot that has to understand language and work in an open and dynamic environment. Therefore, the bottom-up approach is needed. In the bottom-up approach, a robot dynamically learns the connection between words and the real world and

---

[1] ASIMO, http:/www.honda.co.jp/ASIMO

can update this knowledge through interaction with humans. Recently, a field called Cognitive Developmental Robotics (Asada et al., 2001) has been proposed, paving the way for this type of research. Cognitive Developmental Robotics aims to construct an intelligent robot which can behave adaptively by learning through interaction with the environment. The behaviorism approach "Embodied Intelligence", advocated by Brooks (Brooks, 1991), is another bottom-up approach.

For above mentioned reasons, a framework in which the robot makes connections between words and the real world by itself is an efficient way to develop machines that can perform universal tasks and understand natural language in dynamic environments. Thus approaches where robots acquire language developmentally like humans have attracted attention as one possible method to design the robot. Here we would like to suggest such a language acquisition mechanism for a humanoid robot. Our target language is Japanese, and we will use italic when giving Japanese examples.

1.2 Related Work

There are several research groups working on language acquisition for embodied systems. Iwahashi et al. proposed a mechanism whereby a robot arm acquires nouns, verbs, word order and concepts of movements from pairs of movement video and an audio explanation (Iwahashi et al., 2003). In this research, movement was modeled using a hidden Markov model (HMM). The system acquires nouns from audio using statistical learning. Verbs are also acquired by statistical learning and an embedded mechanism that can represent the trajectory of moving object.

Tani et al. described a system using a recurrent neural network (RNN), where a movable arm robot acquires nouns and verbs from pairs of an action pattern and a two word phrase (Tani et al., 2005).

Ogura et al. developed a mechanism where a humanoid robot acquires nouns, verbs, and word order using Self Organizing Incremental Neural Network (SOINN), and understands three words utterances (Ogura et al., 2006).

In the above research, utterances lack the naturalness of natural Japanese language we use, because they do not use particles at all. Fukui et al. experimented with machine language understanding of more natural utterances using an AIBO[2] (Fukui et al., 2004). In their research, the action representation is a row of numbers which indicates an action, such as kicking, heading or making steps, etc. and the AIBO learns words or phrases from pairs of a simple sentence and a pattern of movement.

However, the above mentioned systems do not deal with the understanding of compound sentences containing more than one verb. It is rather obvious that it is much more natural to give orders to robots by using compound sentences such as "go to the next room and bring me a book". Earlier research has concentrated only on spatial concepts such as colors, shapes movements or actions, ignoring the necessity to acquire other concepts, such as time or negation, for understanding compound sentences.

1.3 Our language Acquisition Model

We will develop a language acquisition mechanism that can deal with compound sentences.

---

[2] AIBO, http://www.sony.jp/products/Consumer/aibo/

In a compound sentence, connectives (in Japanese for instance indicating conjunction or negation) join main and subordinate clauses. We try to acquire concepts of time and then ground connectives to them.

Below are explanations of four levels of our language acquisition model based on research on human language acquisition.

(a) First, infants acquire the phonology of their mother tongue. In this acquisition step, infants use categorical perception (Eimas & Miller, 1980), though we are not concerned with phonology acquisition. Assuming that phonology is already acquired and phonologies are connected to characters, we use textual inputs for our system.

(b) Second, the morphology is acquired. Infants learn segmentation of words from continuous speech gradually, using prosody and statistical information (Jusczyk, et al., 1999). Such statistical segmentation has already been developed as a morphological analyzer for textual Japanese such as MeCab[3]. Therefore, we use such an analyzer to overcome the problem of word segmentation in our system. Utilizing this analyzer, our system has knowledge of word segmentation from the beginning.

(c) Third, the words which are segmented are grounded to appropriate meanings in the real world. Nouns are acquired by fast mapping from very few language inputs, because infants ground nouns to their meanings using some innate constraints of cognition (Markman, 1989). Verbs and adjectives have not been studied that thoroughly. Our system does not abstract and acquire nouns, verbs, or adjectives yet.

(d) Finally, we know that infants also infer things inductively (Heit, 2000). Therefore, we believe that infants ground abstract words to meaning, such as connectives that include partial time concept, by inductive inference from actual examples.

Our system acquires connectives by abstracting from pairs of a sentence and an action using the "Three Examples Based Inductive Learning" method which is based on Inductive Learning (Araki & Tochinai, 2000). In this algorithm, we use our system's innate learning ability and innate cognitive ability. The former is the ability to compare if one string contains the other string, and then to parameterize the common part; the latter is the ability to recognize movements based on the final posture of the robot. This learning algorithm is one of original points of this paper, and the acquired connectives have compositionality and can create various new meanings by combining known words or phrases. Moreover, the humanoid robot is taught actions by users using the direct physical feedback. Direct physical feedback means a teaching method where the user teaches the robot by moving its arms, legs, or head directly. Thus teaching method is also original.

2. Suggested Method

2.1 System Overview

Our system is shown in Fig. 1 and it is outlined below.

(a) A user inputs a command in natural language (Japanese) from a keyboard for the humanoid robot. The user can input both a command representing a simple action (see 2.2) and a command representing a compound action (see 2.3).

---

[3] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.jp/

(b) The robot applies rules (see 2.8) or examples (see 2.7) that have been previously acquired, and performs an action. If there is no rule or example which should be applied, the robot does not perform any action.

(c) When the robot performs an action, the user makes the judgment whether the performance is correct or not, and if it is not, the user teaches it the right action (feedback process). After that, the system adds a pair of input command and taught action to the example database.

(d) If the humanoid robot does not perform any action, the user teaches it a proper action by direct physical feedback (see 2.4). Then the system adds a pair of command and action to the example database.

(e) Finally, the system generates rules which represent meanings of connectives from the example database by Three Examples Based Inductive Learning process (see 2.8).
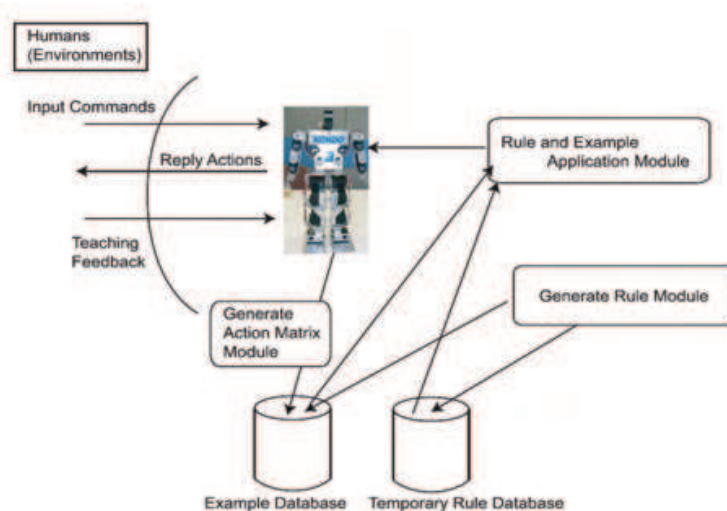

Fig. 1. System overview

2.2 Element of Action

In their work on understanding of order expressions for actions, Shinyama et al. discussed about "the vagueness in instructions and the mistiness in spatial points" (Shinyama et al., 2001). Regarding the vagueness in instructions, robots have to determine whether a user's utterance instructs to perform something or not. Although, for example, "can you raise your hand?" has a form of a question, it may instruct the robot to raise its hand. Concerning the vagueness of spatial points, robots have to determine where the point of the user's instruction is. For example, "raise your hand" has vagueness of height which the hand should be raised to.f

In our previous work (Hasegawa et al., 2007), to make the acquisition of connective simpler, we defined an action as a body movement trajectory of the shortest distance between starting and final positions. The final position which the robot should reach is obtained from user's input. Furthermore, we resolve the vagueness by teaching and averaging. The robot determines a final position of an action as an average point from all inputs taught by all users who input the same command. Within this definition, the system can not deal with actions where the final position changes continuously and intermittently, for example, "to wave a right hand". However, the action is represented as a combination of two actions

which are "to move the right hand to the right side" and "to move the right hand to the left side".

Then, it is necessary to define a minimum unit of action for segmentation of complex movements. We try to perform segmentation based on language inputs from users. Thus, we define that a minimum unit of action is an action which is represented by one sentence containing only one verb. We call it Element of Action (hereafter abbreviated as EoA). We also call a command representing EoA an EoA command. It can be summarized as follows.

(a) EoA shows final physical position of action.

(b) EoA command contains only one verb.

For example, "*migi-te wo agete* (raise your right hand)" or "*migi wo muite* (look to the right)" are EoA commands. This definition of actions has one important problem. The problem is an ignorance of a spatial context dependence of actions. For example, in the input "put your right hand on your left hand", the final position of the right hand depends on the spatial position of the left hand. By our definition of actions, the system can not learn the correct action. However, we regard this as the verb acquisition or the verb understanding problem, because the system has to understand that the verb "put" requires two objects usually. We plan to handle it in our future work. Therefore, in our current system verbs are not analyzed. In the connective acquisition, our system deals with the verbs which require only one object in a limited way.

2.3 Connectives

A connective is defined as follows.

(a) A word which is a conjunction or a conjunction particle.

(b) A segment connecting two sentences, which contains a conjunction or a conjunction particle.

Connectives connect two EoA commands. We also call a command that contains a connective a connective command. For example, "*migi-te wo agete kara migi wo muite* (face to the right after raising your right hand)" is a connective command, and "*kara* (after)" is a connective. According to our definition of connectives, they can contain more than two verbs to connect two EoAs. We define an action which contains more than one verb as a compound action to distinguish it from a single EoA. Compound actions are performed by inputting connective commands. There are not so many types of connectives, but their usage slightly depends on individual interpretations regarding timing, speed or quantity of action. Thus we consider that robots should learn the meanings of connectives from various users. In this case, such knowledge is being acquired like common sense. We use average of all inputs and feedbacks to simulate this process.

2.4 Teaching Method

A humanoid robot learns EoAs and compound actions by being taught by human how to move its body.

There are some methods already developed where a human supervisor teaches actions to humanoid robots. In one of them, a user shows actions to a humanoid robot equipped with vision and the robot imitates "seen" actions (Mataric, 2000) (Schaal, 1999). Another is where a user makes a humanoid robot learn actions from human's motion capture (Nakaoka et al., 2003). However, these methods can teach actions beyond an allowance of a robot's body

structure while our method also helps the user and robot itself to know the functional limits, because its body is not exactly the same as a human's. The robot can not perform an action immediately if it exceeds its physical capabilities. Though there is a way where a user teaches combinations of primitive movements implemented to a robot beforehand, it is difficult to teach new primitive movements and perform corrections.

Therefore, we decided to implement a direct physical feedback method where humans teach actions to a robot by actually moving its body parts. We claim it is a universal and natural method which allows teaching within the limits of any humanoid robot's body structure. This approach also allows robots to reproduce movements with very high certainty. Our direct physical feedback has the merits, but on the other hand the method to teach very complicated actions where many joints must be moved simultaneously and one user is not enough to perform the feedback. Therefore, we do not consider our teaching method as a general teaching method but as an additional method which complements other methods.

The direct physical feedback has an advantage in teaching the haptic actions. For example, in the teaching method through a vision or a motion capture, it is difficult to teach an action where a robot pushes a button or a robot touches something. Because the strength and quantity of such actions is very delicate, our method should be faster and easier to use.

2.5 Humanoid Robot

For our experiments, we used a humanoid robot (KHR2-HV[4]) shown in Fig. 2. The robot is equipped with 17 motors, no sensors and it sends signals describing its motors' state only. We use all 7 motors which are placed in the upper half of the robot's body. The particular motors are abbreviated as follows. [H]:Head, [LS]:Left Shoulder, [LA]:Left Arm, [LH]:Left Hand, [RS]:Right Shoulder, [RA]:Right Arm, [RH]:Right Hand. All motors' 180 degrees movements were divided into 10 ranks, 18 degrees each.
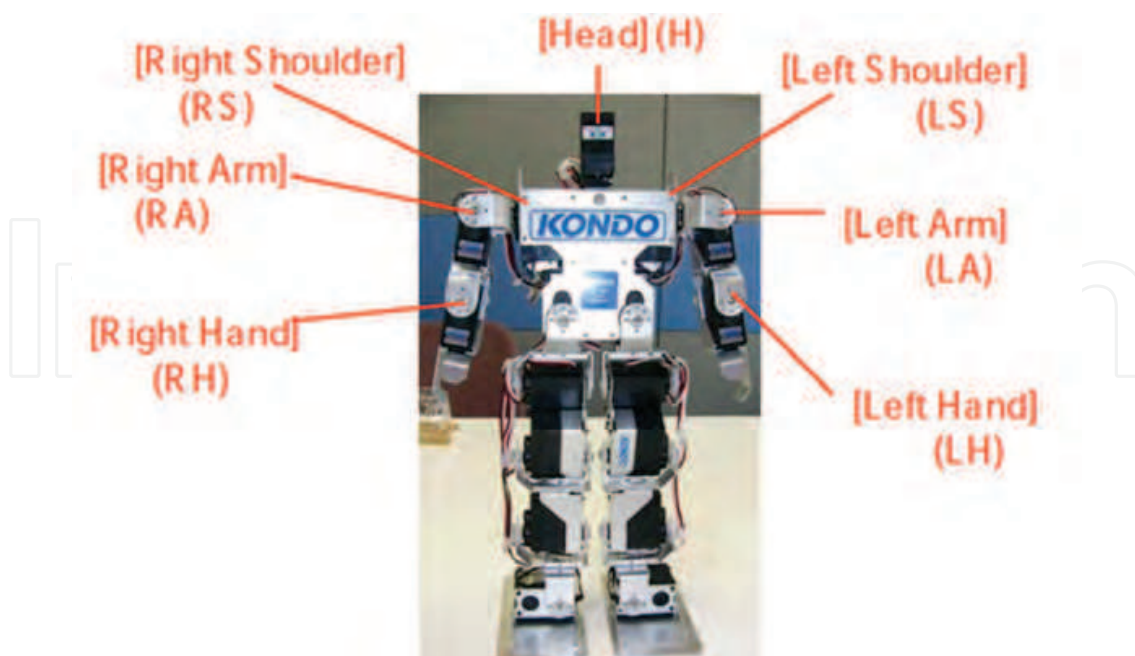

Fig. 2. KHR2-HV

---

[4] Kondo Kagaku Co. Ltd, http://www.kondo-robot.com/

2.6 Representation of Action

An EoA or compound action is represented as a matrix shown in Fig. 3. We call the matrix as an action matrix. If a user moves the robot's body, then it measures all the angles of motors in degrees and digitalizes them every 1 second. A row of an action matrix corresponds to 1 second, and each row corresponds to every motor's state in this particular second. If a motor does not move, degree of its angle is described as "x". An action matrix has 28 rows and 8 columns, but the number of lines depends on hardware's restrictions. To make the acquisition of connectives simpler we do not use motor speed information yet but we plan to use it when acquiring adverbs. Therefore the row of speed is still not used and the matrix can not represent EoAs containing changing speed, for example "*migi-te wo hayaku agete* (raise your hand quickly)".

	[H]	[LS]	[LA]	[LH]	[RS]	[RA]	[RH]	[Speed]
1sec	x	0	1	2	x	x	x	x
2sec	x	5	4	3	x	x	x	x
3sec	x	9	x	x	7	x	x	x
28sec	x	x	x	5	2	x	x	x

Fig. 3. Action matrix

2.7 Example

The system makes examples from actual inputs from users. An example is a pair of a natural language part and an action matrix part. A natural language part is an EoA command or connective command. EoA commands and connective commands input by users are segmented into morphological elements first using MeCab. Then the system changes ending of verbs (which vary depending on conjugation) to original forms by a morphological analyzer. We perform morphological analysis to make learning more effective. It means that analyzing morphological elements and absorbing a variety of changing verb endings will be accomplished only by increasing the number of inputs. Therefore, we use a morphological analyzer instead of a vast amount of inputs. An action matrix is generated automatically based on its correspondence to the natural language part. The correspondence between those parts is being learned from users' input and feedback.

3. Generation of Rule by Three Examples Based Inductive Learning

3.1 Distinction Examples of EoAs and Compound Actions

All inputs from users are stored in an example database. Therefore, the system needs to distinguish examples of EoAs and compound actions (Fig. 4). Firstly, the system brings out any three examples from the example database. In natural language part, if one example's string includes the other two example's string, then the system distinguishes such example as an example of compound action, and the other two examples as an example of EoAs. Next, the system extracts EoA example's final state of motors from action matrix part (Fig.

5). Because there are some examples which are the same language part and different final state of motors, their final states are averaged.
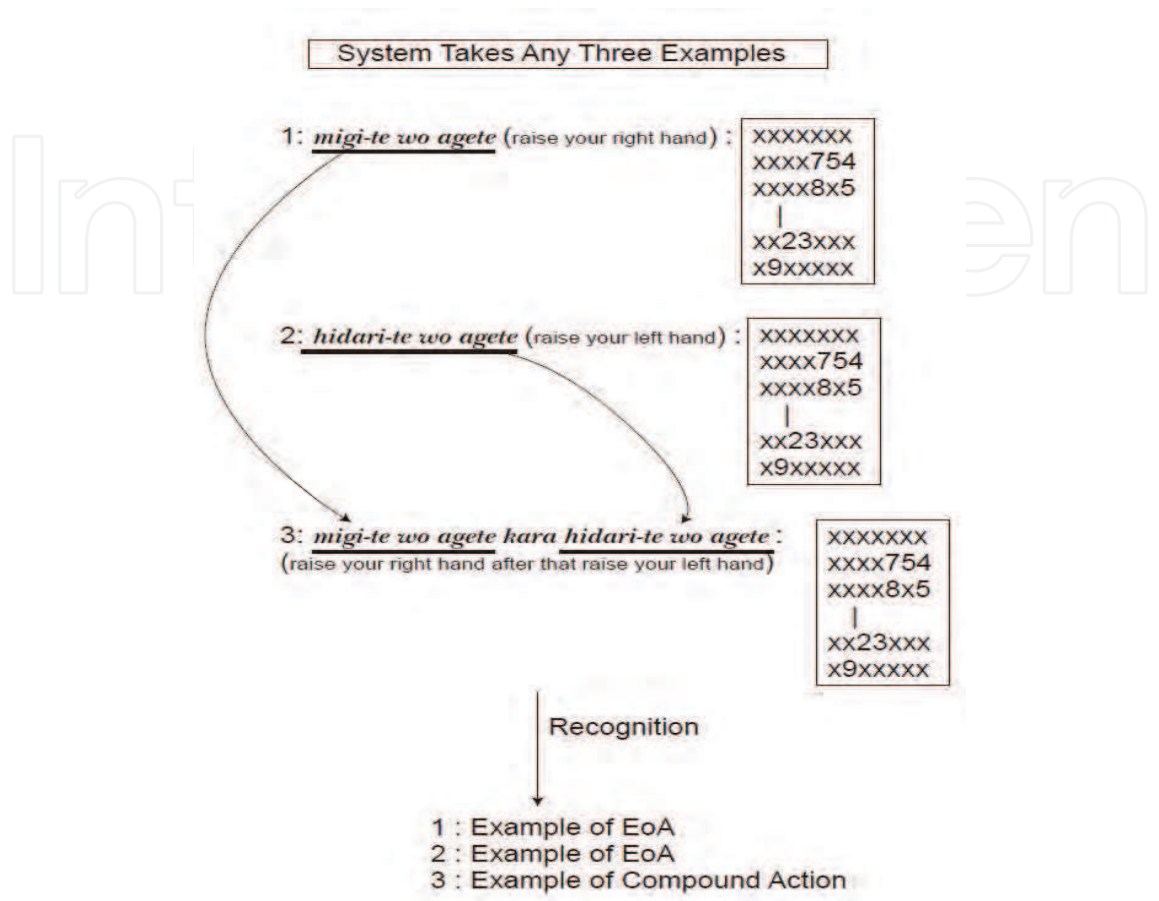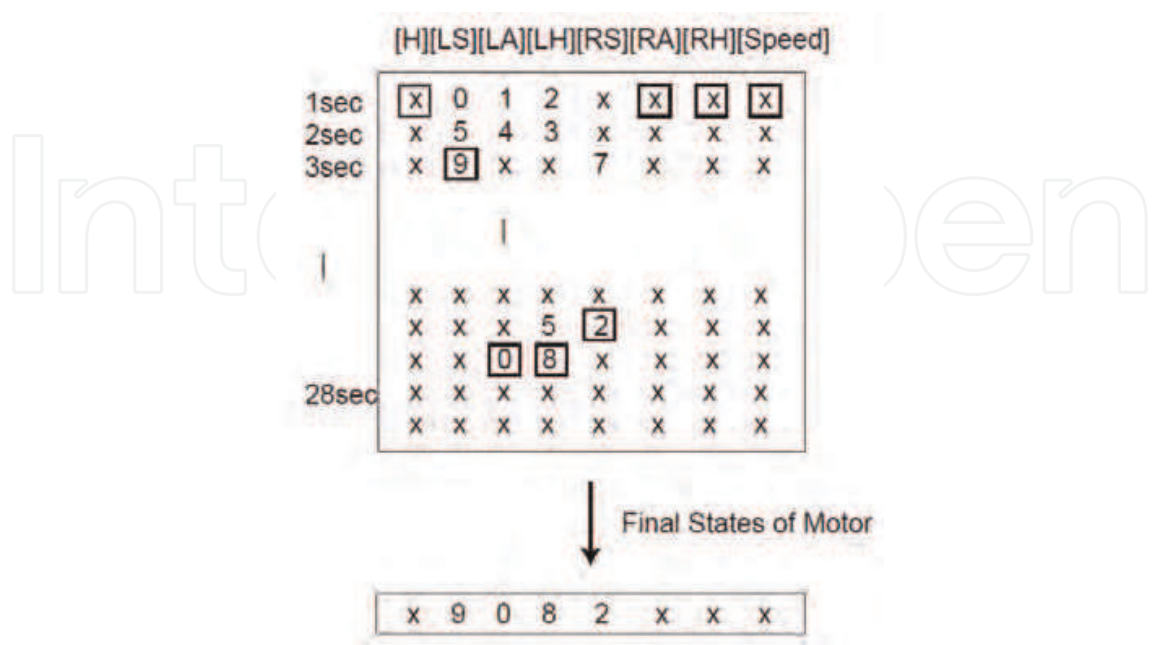

Fig. 4. Distinction of EoA and compound action

Fig. 5. Final states of motor

3.2 Abstraction of Meanings of Connectives by Three Examples Based Inductive Learning

Meanings of connectives are abstracted from examples of compound actions which contain connectives (Fig. 6). In this paper, as our targets are compound sentences which have one connective, we do not implement recursiveness of Inductive Learning in our method. Therefore, the abstraction of the method is performed one time.

(a) The system brings out any example of compound action containing two EoAs.

(b) In natural language part, the system parameterizes substrings of EoA commands of the connective command string as @1 and @2. Then, the remaining part of the string is assumed to be a connective.

(c) Next, in the action matrix part of compound action, the system finds rows which accomplished final positions of EoAs, and parameterizes the rows as @1 and @2. Then, the structure of the remaining matrix represents the meaning of the connective.

(d) Lastly, a rule becomes a pair of a parameterized natural language part and a parameterized action matrix part.

In a parameterized action matrix, the structure, for example the sequence of parameters or the time delay from the first parameter to the next parameter of action matrix, represents meaning of connective. Originally Inductive Learning makes abstraction reflexively by abstraction among rules. However, in our system, Three Examples Based Inductive Learning makes one abstraction for the first step. Therefore, we can deal only with two EoAs combinations.
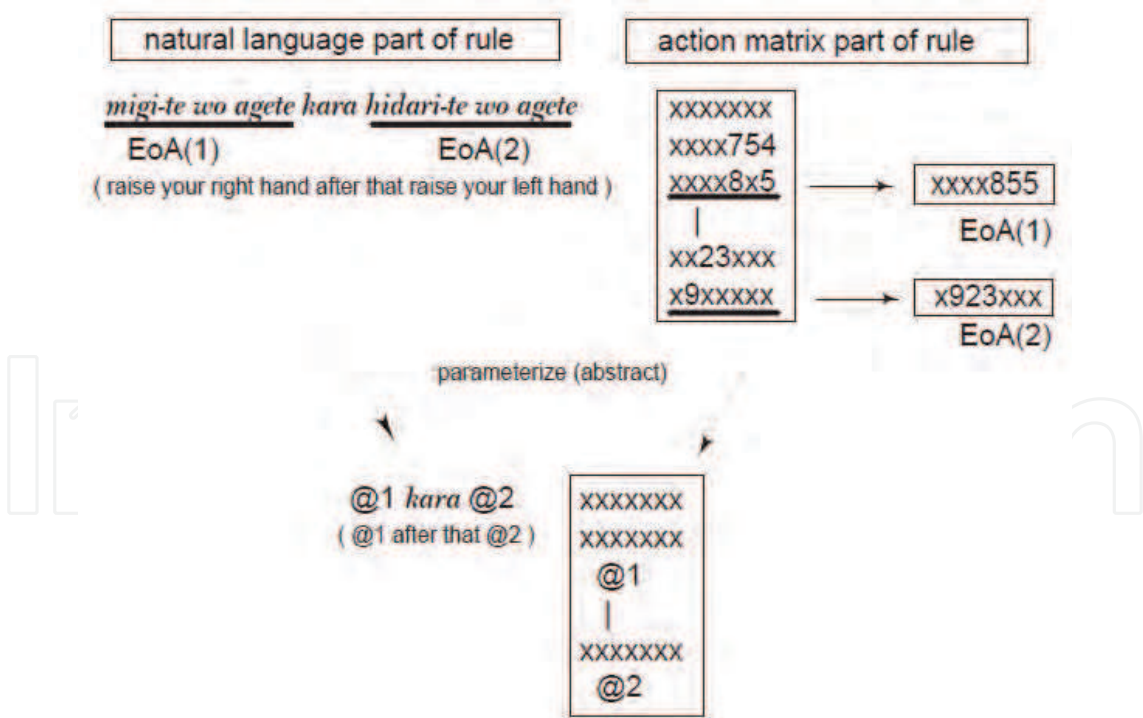

Fig. 6. Gengeration of rules by Three Examples Based Inductive Learning

3.3 Averaging of Rules

Rules are classified into 6 structures as follows.

(a) There is no parameter.
(b) There is only parameter @1.
(c) There is only parameter @2.
(d) There are @1 and @2 in normal ascending order.
(e) There are @1 and @2 in reverse ascending order.
(f) There are @1 and @2 in the same raw.

There are rules about a connective having different structure of action matrix, because the users' feedbacks slightly differ for actions with the same connectives. Therefore, the system has to average the examples contained in each structure. Firstly, the system distinguishes each rule about one connective as 6 structures. Then, in each structure, positions of parameter rows are averaged.

3.4 Adaptation Value of Rules

When there are averaged rules which have different structure, the system needs a criterion to select the best one. For this reason, the system calculates adaptation values of rules. The system counts the number of rules which have identical structure. Then the system regards the number as the adaptation value of the rule. Because the more number of teachings of the same rule's structure increases the higher the credibility of the rule becomes, a rule which has the largest adaptation value is referred if there are more rules which can apply.

3.5 Application of Examples and Rules

When a user makes a command in natural language, the system tries to perform the given action by applying previously taught examples and abstracted rules. The process is outlined below.

(a) Firstly, the system tries to apply rules. If input commands contain a connective string inside the rules database, then a rule which have the highest adaptive value is chosen.

(b) Then, the system chooses two EoA examples for parameters in natural language part from the rules database. Following the structure of the action matrix of the rule, final states of motors of the two EoA examples are inserted in parameters and the action matrix is performed by the robot.

(c) Secondly, if the system can not apply any rules, the system tries to perform the input command by applying actions from previous examples. Then, an example which has the same string as the input command is chosen and the robot performs an action based on the example's action matrix corresponding to its equivalent in language part.

(d) Finally, the robot performs no action if there is no rule or example which can be applied to the input command. Because the meanings of connectives are abstracted, the system can deal with unknown combinations of EoAs. There is an uncountable number of EoAs. Their combinations are a product of the EoAs number and EoAs number minus one, however it is not necessary to teach a robot this amount of combinations when using our method.

4. Experiment

We implemented the system and experimented on learning connectives. Learning experiment and evaluation experiment were conducted.

4.1 Learning Experiment

In this experiment, we determine if the learning system works and the learning process converges. We made four participants (age 20-30, all male students of graduate school majoring in science) input commands and taught actions to the robot. The system accumulates previously input knowledge without deleting it when users change. The flow of the learning experiment is shown in Fig. 7.

1. A participant inputs a command in natural language (Japanese) from a keyboard for the humanoid robot. A participant can input one of ten designated EoA commands or one connective command containing two of those ten EoAs and any connective he wants. Then all connectives are free to choose by participants to make this connective learning experiment fair.

2. The robot applies rules or examples that have been previously acquired, and performs an action. If there is no rule and no example which could be applied, the robot does not perform any action.

3. In the end the user evaluates the robot's performance by using marks shown below and teaches the robot the action when needed.

(a) Correct Response: The system's reply is regarded by participant as a correct movement.

(b) Semi-correct Response: The system's reply is close to the correct response but is not perfect in the participant's opinion. We define correct response and semi-correct response as proper responses.

(c) Erroneous Response: Participant regards the system's reply as incorrect and teaches the correct action to the robot.

(d) No Response: There is no response from the robot due to the lack of particular connective rule(s) and previous teaching of the EoA. The participant has to teach the robot a correct action.

4. Repeat them.

The system has no knowledge in the beginning, therefore all EoAs are taught by participants. In this situation, thing we have to pay attention to is the variety of language. For example, "raise your right hand" and "rise up your right hand" mean the same action. Because there are many ways how to represent the same action, users have to teach a vast number of EoAs to the robot. This problem can probably be solved by paraphrasing or other techniques; however we regard the problem as one of the future works. Therefore, for this stage we restricted the number of EoAs to 10 in order to make learning of connectives more effective. We prepared in advance a questionnaire which was answered by 10 participants. The questionnaire asked what action the participant would like this robot to perform. From this questionnaire results we collected 100 actions, and we chose randomly 10 EoAs which our robot can perform with its upper body. We used the 10 following EoAs.

(a) *migi-wo muite* (look to the right)

(b) *hidari-wo muite* (look to the left)

(c) *migite-wo atama-ni oite* (put the right hand on your head)

(d) *hidarite-wo koshi-ni oite* (put the left hand on your waist)

(e) *Ain shite* (perform the Ain3)

(f) *komanechi shite* (perform the Komanecchi4)

(g) *keirei shite* (salute [me])

(h) *akushu shite* (shake hands [with me])

(i) *gattsu pozu shite* (clench [hold up] fists in triumph)

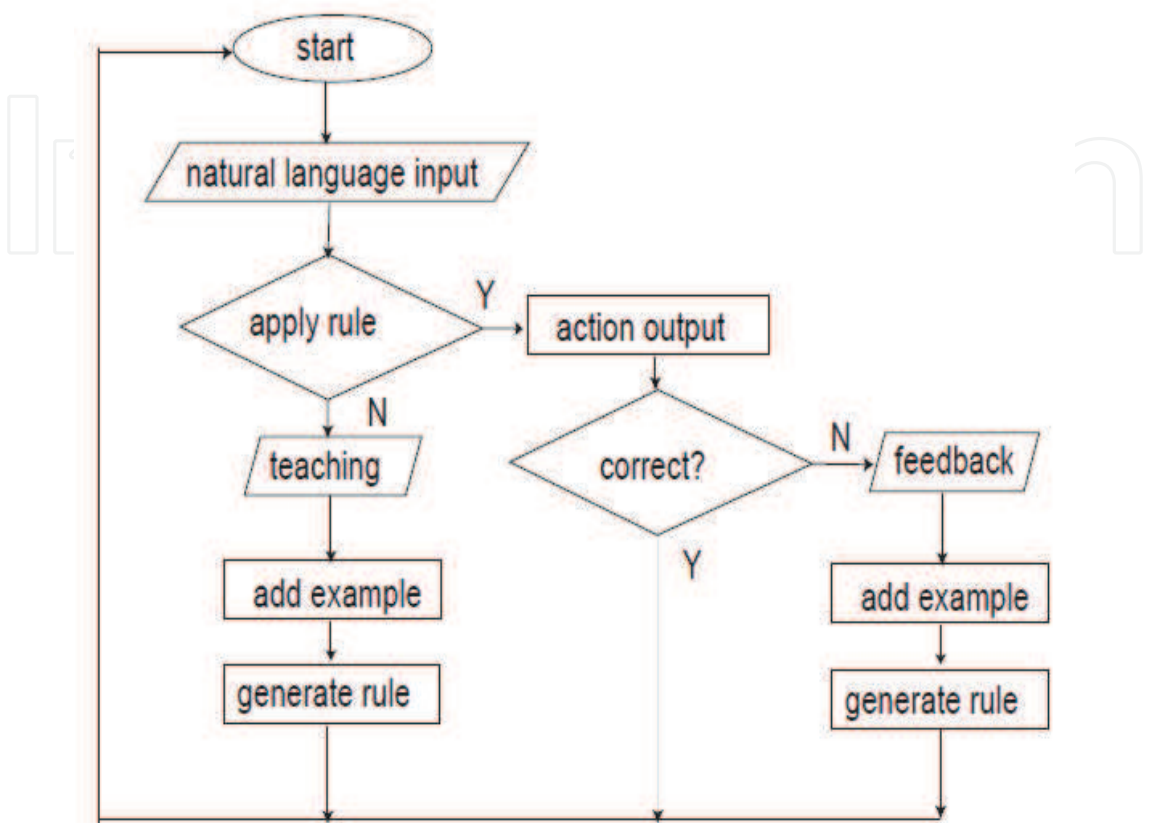(j) *hidarite-wo mae-ni dashite* (hold out your left arm)

Fig. 7. System flowchart

4.2 Result of Learning Experiment

The participants input a total number of 210 commands which contained 157 connective commands (all 157 connective commands differed). The number of connectives actually used by participants was 25 (see Table 1). The process of learning experiment is shown in Fig. 8 and 9. Fig. 8 shows accumulative shifts of proper response ratio, error response ratio and no response ratio of all responses. The proper response line shows gradual increase, while the no response line shows a radical decrease in the beginning and a gradual decrease later. This proves that our system kept learning EoAs and connectives from the state of no knowledge, therefore it gradually made the robot perform proper movements. We can also observe that the system can correct wrong rules, as the erroneous response line shows a slight, gradual decrease. It proves that the system could adapt to changing participants (four of them in this case) and could change rules correctly due to the feedback process. Fig. 9 shows a shift per 10 commands of proper response ratio to 157 connectives commands excluding EoA commands. The proper response line comparatively gets steady after inputting about 80 connective commands. It means, though the system had no enough knowledge to reply with proper responses in the beginning, the robot has corrected its knowledge enough to perform proper movements later.

-kara (and then) **
nagara (and at the same time) **
mae-ni (before) **
ato (after) **
-naide (without) **
– (–) **
-to doji-ni (while) **
-kara yukkuri (and after that slowly) *
sugu (do just after) *
-to misekakete (pretend to do and after that) *
ato-de (after that)
sukoshi[Kanji] matte-kara (after waiting a while)
ato-de takusan matsu (wait a few minutes and after that)
tsutsu (and)
shibaraku shite-kara (after some time)
sono mama (keep on)
ato-de (Kanji) (after that)
soshite (after)
soshite takusan matte-kara (after waiting couple of minutes ...)
toki (when ...)
kara sugu (soon after)
sukoshi [Kana] matte-kara (after waiting a second)
takusan matte-kara (after waiting several seconds)
-to omowasete (after making one think that you do...)
... furi-shite (pretend to ... and after that)

Table 1. Connectives from participants
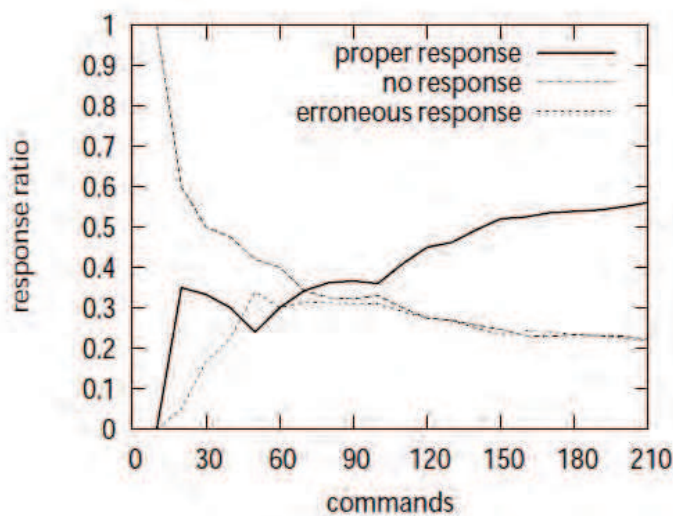
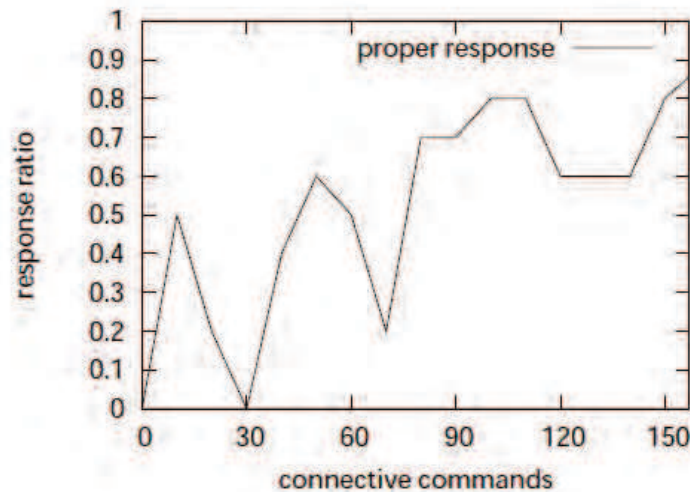Fig. 8. Shift of the system response

Fig. 9. Shift of the system response (connective commands)

4.3 Evaluation Experiment

In this experiment, we determine if rules acquired in the learning experiment are correct. We asked four participants to evaluate the rules. The participants (age 20-30, all male and majoring in science) were different from the learning experiment participants. 25 connectives were proposed by participants during the learning experiment. However, there were cases where participants made mistakes while teaching. We took them into consideration and chose connectives input by participants more than three times as target rules of this evaluation. There were ten of such rules obtained (rules marked "**" or "*" in Table 1).

By adding two EoAs to each of ten target connectives, we made ten connective commands. Then the robot performed those connective commands, and the four participants were shown the commands and the actions. The participants evaluated actions with the following evaluation values.

(a) Success: Score=3
(b) Almost Success: Score=2
(c) Failure: Score=1

4.4 Result of Evaluation Experiment

We define a "correct rule" as a rule which was evaluated as "Success: Score=3" or "Almost Success: Score=2" by all participants. The overall precision was calculated by Formula (1), and the precision of connectives acquisition was 0.7.

From the evaluation results we can find that the system can acquire connectives which signify simple relation of order conjunction (for instance, "*nagara* [and at the same time]" or "*mae-ni* [before]"). We partially acquired a concept of time using the embodied system. Furthermore the system acquired a connective that represents meaning of logical negation. That is "*naide* (without)". The system could not acquire three connectives. First of them, "-*kara yukkuri* (and after that slowly)" was scored 1 by only one participant. This connective contains an adverb. However, other participant scored 3 for the same performance. It shows that the evaluation of performance requested with the same command depends on

participants. In the second case, "*sugu* (just before)" was scored only 2 or 1. This connective also contains an adverb. The last one, "*misekakete* (pretend to do and after that)" is scored 1 by all participants. It contains a verb and the system still can not understand verbs correctly. The robot still can not represent verb meaning. Therefore, the connectives including verbs are not acquired. We plan to work on a mechanism to acquire verbs in the near future.

$$\text{Precision} = \frac{\text{the number of correct rules}}{\text{the number of rules applied three or more times}} \quad (1)$$

Rule	Evaluation (participants A, B, C, D)	sum
@1 *kara* @2 (@1 and then @2)	3, 3, 2, 2	10
@1 *nagara* @2 (@1 and at the same time @2)	3, 2, 3, 3	11
@1 *mae-ni* @2 (@2 before @12)	3, 3, 2, 3	11
@1 *ato* @2 (@2 after @1)	3, 3, 2, 3	11
@1 *naide* @2 (@2 without @1)	3, 3, 3, 3	12
@1 @2 (@1 @2)	3, 3, 3, 3	12
@1 *-to douji-ni* @2 (@1 while @2)	3, 2, 2, 3	10
@1 *-kara yukkuri* @2 (@1 and after that slowly @2)	1, 2, 2, 3	8
@1 *sugu* @2 (@1 just before @2)	2, 2, 1, 2	7
@1 *-to misekakete* @2 (pretend to @1 and after that @2)	1, 1, 1, 1	4

Table 2. Evaluation

5. Discussion

5.1 Compositionality of Rules

In Fig. 9, the system responded to unknown connective commands with an accuracy of about 0.7 after 90 commands. These commands are new combinations of EoAs. The number of possible commands that can be made by such combinations is 2,500, because the number of EoAs is ten and the number of connectives that were input by users is 25. Still the system could respond with high accuracy by learning only a small number of commands. Thus, we can conclude that the acquired connectives that are abstract rules have compositionality and the rules can handle many new meanings by combination of simple sentences. Such compositionality is one important aspect of natural language. The system is able to produce different (and not previously input) results by varying EoAs. This is one advantage of our learning algorithm.

5.2 Learning Algorithm

In the learning algorithm which we suggested, the system can only connect two simple

sentences, though humans can create complex orders with many clauses following each other, for example "@A and after that @B and then @C and after that ...". This is because the process of comparing and parameterization is not recursive. We have to extend the learning algorithm to compare not only examples, but also rules recursively.

Furthermore, we utilized an innate learning ability and an innate cognitive ability as a basis for the learning algorithm. The cognitive ability is the ability to recognize movements based on the final posture of the robot. However, this is too simplistic to be able to recognize many types of movements. With this limited basic ability, the system cannot correctly recognize for example waving movements or movements in which the final posture depends to the context. Therefore, we have to extend the cognition ability.

5.3 The Number of EoAs

In this paper, we do not abstract the EoAs themselves. Therefore, the number of EoAs grows large. If one user inputs a command expressed a little bit differently than the other users' commands, then the system recognizes it as a completely different input and does not perform any action. The users may have to teach EoAs endlessly. For example, we have to teach "raise your right hand" and "raise your left hand" separately, because the commands mean different actions. However, if the system acquired the meaning of the verb "raise" after being taught "raise" a few times, then the system could autonomously create the unknown action represented as "raise your left hand" by combining it with meaning of "left hand". To solve this problem, we need acquisition of verbs, adverbs and nouns. To acquire the meanings of verbs, we need a representation system for them. Such system has to be able to handle both abstraction and symbol manipulation.

5.4 Two Types of Ambiguity of Connectives

We believe that there are two types of ambiguity in connectives. One is ambiguity depending on a user's peculiarities and the other is ambiguity depending on other context factors where the connective usage patterns may change for the same user. The former ambiguity can be resolved by distinguishing users, though we do not currently implement that. However, the latter ambiguity remains a problem to be solved. For example, "raise your right hand before raising your left hand" is ambiguous because the robot cannot decide if the right hand should be lowered before raising left hand or the right hand should remain raised.

6. Conclusions and Future Work

We proposed a language acquisition model based on an Inductive Learning Language Acquisition Model. Then we focused on connectives acquisition under the assumption that our system already acquired phonology and morphology. The proposed connectives acquisition algorithm is that the system inductively abstracts meanings of connectives from pairs of connectives and motor patterns. As a result of two experiments, we confirmed that the system learns connectives with an accuracy of 0.7 and the acquired connectives have compositionality.

In our future work, we will concentrate on improving the recursive learning algorithm. After that, we will work on a verb acquisition algorithm for humanoid robots.

7. References

Araki, K. & Tochinai, K. (2001). Effectiveness of Natural Language Processing Method Using
Inductive Learning, *Proceedings of the IASTED International Conference on Artificial
Intelligence and Soft Computing*, 0-88986-264-8, Cancun , pp.295-300

Asada, M.; MacDorman, K.; Ishiguro, H. & Kuniyoshi, Y. (2001). Cognitive developmental
robotics as a new paradigm for the design of humanoid robots, *Journal of Robotics
and Autonomous System*, Vol.37, pp.185-193, 0262-8856

Brooks, R. A. (1991). Intelligence without representation, *Artificial Intelligence*, Vol.47,
pp.139-159, 0004-3702

Eimas, P. D. & Miller, J. L. (1980). Contextual effects in infant speech perception, *Science*, 209,
pp.159-207, 0036-8075

Fukui, H.; Araki, K. & Tochinai, K. (2004). The Method of Acquisition of Corresponding
Relation between Language and Act Using Inductive Learning with Genetic
Algorithm (in Japanease), *Journal of Infomation Processing Society of Japan*, Vol.45,
No.10, pp.2428-2437, 0387-5806

Hasegawa, D.; Rzepka, R. & Araki, K. (2007). Evaluation of Connectives Acquisition in a
Humanoid Robot Using Direct Physical Feedback, *Advances in Artificial Intelligence*,
Vol.4830, pp.664-668, 978-3-540-76926-2, December 2007, Springer, Surfers Paradise
Australia

Heit, E. (2000). Properties of Inductive Reasoning, *Psychonomic Bulletin & Review*, Vol.7,
pp.569-592, 1069-9384

Iwahashi, N. (2003). Language acquisition by robots? towards a new paradigm of language
processing, *Journal of Japanese Society for Artificial Intelligence*, Vol.48, No.1, pp.49-58,
1346-0714

Jusczyk, P. W.; Houston, D. M. & Newsome, M. (1999). The beginnings of word
segmentation in English-learning infants, *Cognitive Psychology*, 39, pp.159-207, 0010-
0285

Markman, E. (1989). Categorization and naming in children: Problems of induction,
Cambridge, Mass.: The MIT Press, 0-26263-136-9

Mataric, M. J. (2000) Getting Humanoids to Move and Imitate, *IEEE Intelligent Systems*, v.15
n.4, p.18-24, 0-7803-9731-7

Nakaoka, S.; Nakazawa, A.; Yokoi, K. & Ikeuchi, K. (2003). Symbolic Motion Description for
a Dancing Humanoid Robot (in Japanease), *Technical Report of IEICE, PRMU2003-
137*, 0913-5685

Ogura, T.; He, X.; Satou, A. & Hasegawa, O. (2006). Developmental Word Acquisition And
Grammer Learning by Humanoid Robots through SOINN (in Japanease), *IEICE
Technical Report PRMU2006-100*, pp.51-56

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and
perception, *Artificial Intelligence*, Vol.167, pp.170-205, 0004-3702

Schaal, S. (1999). Is Imitation Learning the Route to Humanoid Robots?, *Trends in Cognitive
Science*, Vol.3, No.6, pp.233-242, 1364-6613

Shinyama, Y.; Tokunaga, T. & Tanaka, H. (2001). "kairai" - Software Robots Understanding
Natural Language (in Japanease), *Journal of Infomation Processing Society of Japan*,
Vol.42, No.6, pp.1359-1367, 0387-5806

Sugita, Y. & Tani, J. (2005). Learning samentic combinatoriality from the interaction between linguistic and behavioral processes, *Adaptive Behavior*, Vol.13, No.1, pp.33-52, 1059-7123

Winograd, T. (1972). Understanding Natural Language, *Academic Press*, 0127597506, NY
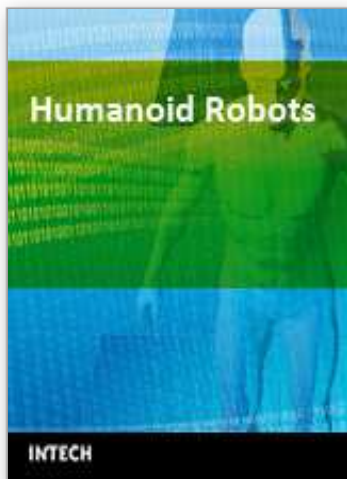
Humanoid Robots

Edited by Ben Choi

Humanoid robots are developed to use the infrastructures designed for humans, to ease the interactions with humans, and to help the integrations into human societies. The developments of humanoid robots proceed from building individual robots to establishing societies of robots working alongside with humans. This book addresses the problems of constructing a humanoid body and mind from generating walk patterns and balance maintenance to encoding and specifying humanoid motions and the control of eye and head movements for focusing attention on moving objects. It provides methods for learning motor skills and for language acquisition and describes how to generate facial movements for expressing various emotions and provides methods for decision making and planning. This book discusses the leading researches and challenges in building humanoid robots in order to prepare for the near future when human societies will be advanced by using humanoid robots.

INTECH
open science | open minds