Studies on Discourse Understanding Method and Cognitive Semantic Model

Shun Ishizaki, Shigenori Tanaka, and Mutsumi Imai
Environmental Information Faculty,
Keio University
5322 Endoh, Fujisawa
JAPAN, 252
e-mail: ishizaki@sfc.keio.ac.jp

1 Introduction

In our daily life we easily invent grammatically incorrect sentenses by hesitating or correcting ourselves. The traditional, grammar and syntax based ways of natural language processing or machine dictionaries were insufficient to interpretate such sentenses, even though they all do make sense and we human beings have no problem of understanding them. In this thesis we propose you to divide a dialog into smaller chunks of meanings, as it is rather a chain of fragments than that of sentenses. We communicate each other by exchanging such meaning chunks.

This paper includes three researches in order to get further information on human cognition to make our theory feasible. First is a theoretical consideration of cognitive activities form the sense-making view points. We prepared and edited a corpus of conversational language (English) consisting of 50,000 running words. Based on the corpus, two discourse markers 'you know' and 'I mean' were statistically examined, then 'because' was examined to explore its usage in the corpus. In the third year, sense-making theory was built base on such data.

Next one is on the affinity between nouns and verbs by using SD method using 35 sets of diametrical adjectives. We selected 12 Japanese verbs which correspond to an English word 'to break' and 80 various concrete nouns as their objects. We then examined the semantic acceptability of each combination of a verb and a noun. We did the same thing with abstract verbs.

Finally, a metaphor analysis model is proposed to get salient features between two words such as 'T is V' type metaphors. We did a cognitive experiment with 20 subjects showing 35 sets of diametrical features, such as big and small, fast-slow. We made an connectionism-applied analyses system called COFFS using the result. A concept association experiment was done to construct concept dictionary similar to the human memory structure, then a flexible information retrieval method is proposed by using the concept dictionary.

2 Sense-making Theory

The goal of this three-year long study was to develop a theoretical framework to adequately describe a coversational dialogue from the perspective of sense-making. In the first year, we prepared and edited a corpus of conversational language (English) consisting of 50,000 running words. The corpus was put into a computer and a computer program was developed to facilitate the data analysis.

On the basis of the corpus (CHATDATA), we made an attempt at describing the formal characteristics of conversational language in comparison with literary language, and as a result, we identified 'fragmentariness' and 'collaboration'. We paraphrased the notion of

fragmentariness as 'chunk' (the semantic unit) in order to advance a linguistic analysis of CHATDATA. Two discourse markers, 'you know' and 'I mean, were then selected in a case study (Case study A). Location and frequency of occurrence of these two markers were first statistically examined, and then, the functional roles of the markers were discussed with reference to the process of sense-making. We suggested that 'you know' is an Other-orientedmarker with 'I mean' being a Self-oriented marker.

In the second year, CHATDATA was refined in terms of accuracy along with the elaboration of the computer program. The nature of conversa- tional language was further explored in Case Study B, which focussed on 'because;, an introducer of subordiante clause. Here again, a stat- istical analysis was made on location and frequency of occurrence of 'because' and other subordiante markers in CHATDATA. The word 'because' predominantly appeared clause-finally. This showed a sharp contrast with 'if', which strongly preferred the clause-initial position. The conjunction 'when' appeared both clause-initially and clause-finally. In order to see the functional role of 'because', a semantically similar word since was compared as a reference. We suggested that 'because' is typically used to set a discourse space in order to justify the pre- vious utterance. We also suggested that the grammatical treatment of 'because' as a subordiante conjunction is highly questionable.

In the final year, an attempt was made to develop a theory of sense- making, which differs markedly from a semantic theory in several respects. To mention one important difference, a semantic theory accepts the assumption that a word has meaning, and sets its goal of describing the structure of meaning (i.e., multiple senses), as exemplified in the Lexical Network Model.

There are two types of shared orderness operative in the process of sense-making: orderness belonging to the way sense-making is carried out, and orderderness belonging to the semantic content of a signifier. In the case of the Japanese language, particles have the key function (the operation function) of indicating how semantic chunks are put together to match the interpreted context. In addition, verbs have the schema function of allocating semantic cases to nouns so as to describe an event. The semantic content of a signifier should be shared among speakers of the same language to a certain extent. We account for this shared orderness in terms of the process of concept formation (or categorization and conceptualization). Three cognitive activities are assumed in a theory of sense-making: differenciation, generalization, and typicalization. Typicalization is a key to under- standing shared orderness belonging to the semantic content of a signifier.

3 Affinity between nouns and verbs

We have selected a group of Japanese verbs in the semantic domain of 'breaking' to study the mechanism of contextual meaning shift in verb polysemy. Our first attempt (Iwata et al., 1994) was to examine whether and how verb meanings are shifted around when they take different nouns. We selected 5 representative Japanese verbs in the domain (oru, mageru, kiru, yaburu, waru). We also selected 5 concrete nouns and 5 abstract nouns. We constructed 50 sentences by systematically inseterting each of the 10 nouns in the object position for each verb. Two experiments were conducted with the material.

In the first experiment, we asked college students to rate each of the sentence on the 1-5 scale in three respects: (1) metaphoricity (i.e., if the verb meaning is used metaphorically in the given sentence); (2)typicality (i.e., if the verb meaning in the given sentence is typical for the verb); and (3)naturalness (if the verb meaning in the sentence is natural and acceptable). Second, using the Semantic Differential Method, we presented the sub-

jects with a list of 18 adjective pairs that seemed relevant for describing verb meanings for these verbs. The example adjective pairs included "abstract-concrete", "intentional-non-intentional", "desctructive- nondesctructive", "momental-continuous". For each of the 18 paris we had the subject rate the degree of relevance of the given feature(i.e., the adjective pair) on the -3 to 3 scale. This procedure was repeated for all 50 sentences. In addition, the subject rated each of the 5 verbs by themselves with no context. This was done to examine the default feature-weighting patterns for these verbs.

The results of the two experiments lead us to conclude that (1) nouns (especially those that are assigned the Theme role) play an extremely important role in shifting verb meanings and that (2) each verb has a preference for certain type of nouns for its object (i.e., the Theme argument) so that properties of the preferred nouns should be incorporated in the verb's core/prototypical meaning.

We conducted another series of experiments(Iwata et al., 1995). The first experiment was designed so that we can group nouns denoting concrete objects with respect to how they interact with the various verbs in the semantic domain of 'breaking'. We selected 12 verbs (kowasu, mageru, kudaku, oreru, etc.) from the target domain. We also selected 80 nouns denoting concrete objects so that they represent a wide range of physical objects with various functional and physical properties such as size, shape, solidity and rigidity, etc. For each verb, we constructed 80 different sentences with each of the 80 nouns serving as the object of the verb. Thus, a total of 960 sentences were constructed. College students were asked to rate the naturalness of the verb's usage in each sentence on the scale of 1(very unnatural) to 5(highly natural). A resulting 80X12 matrix with a mean naturalness score for each item was submitted to a cluster analysis.

Based on the results of the analysis, we looked at commonalties among nouns within a group and differences that distinguish one group from another. Through this process, we extracted four components that seemed important to characterize the within group commonalties and between group differences. The four components are: Individuality, Functional complexity, Solidity/Rigidity, and Dimensionality of shape.

4 Semantic analysis system for metaphor expression : COFFS

Conventional concept dictionaries include hierarchical information such as higher-level concepts and lower-level concepts, and constraint information such as relations with other concepts and slot-filling information(Ishizaki, 1990).

We introduce new conceptual information to the concept dictionary because the conventional type is not enough to describe metaphorical expressions. One of them is semantic metric 3D space using Multi-dimensional Scaling method into the set of concepts. The other is to describe the concept meaning precisely by giving a bundle of many types of features to each concept and then treat the dynamics of the concept meaning change quantitatively.

This section proposes a new method named COFFS(COmmon Features Finding System) which extracts common salient features between Topic concept and Vehicle concept in a context such as 'T is V' using a connectionist model.

We did a cognitive experiment with 20 subjects showing 35 sets of diametrical features, such as big-small, fast-slow. The 35 sets are used for all of the 66 concpets. In the first experiment, the concepts are shown independently to the subjects. In the second, the concepts are shown in context such as 'concept1 is concept2' where concept1 is topic and concept2 is vehicle. The diametric features are classified into 7 steps. The subject was requested to decide which step is appropriate for the concept shown in the figure at the

display. The result in the first experiment is used for the parameters of f's in the figure1. The number of f's are 35 and T and V corespond to the Topic and the Vehicle.

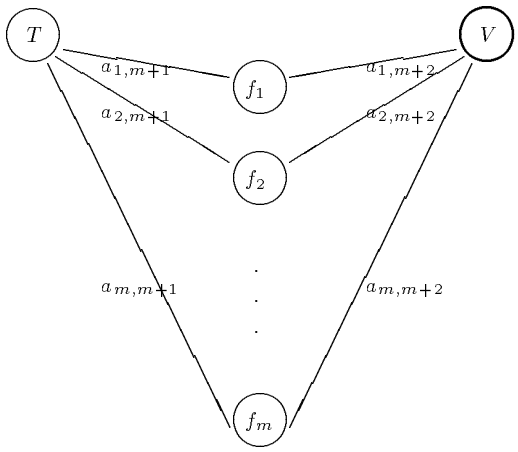

Figure 1 COFFS Structure

The formulation of this model is as follows.

x(t+1) = A * x(t) + c (t=0,1,2,...) (1) x(0) = c (2)

where x(t)(i=1,2,...m+2, m=35) is a vector for nodes of f's, T and V at the time 't' to show its activity value. The bigger the value, bigger its salincy. 'c' is a stationary input to the system. The ith element of c corresponds to a degree of similarity between the ith element of T's and V's f. The element values of the transition matrix A is decided by using the first experiment.

The 5 biggest values for each metaphor sentence are added and shown in the Table1. The better metaphorical expressions are listed at the top positions and the worse ones are listed at the bottom positions, where the salient 5 features are selected among 35 features.

Table1 Example of Metaphor Analysis (Typical 6 examples among 12 ones)

Examples T is V	Active Features 5 Biggest	5 Biggest TOTAL	MDS Distance
Hacker is a magician	skillful,applied,matured, artificial, difficult	1.25	1.09
Computer is a slave	heavy,dark,artificial, many,cold	1.11	1.55
Traffic jam is the sea	big,deep,many(much), continuous,successive	1.07	1.50
Technology is an embryo	complex,new,noble, good,real	0.97	1.79
Inventor is a pioneer	difficult,good, wide,humane,strong	0.88	0.41
Expert is an alien	strong,difficult, smart,cold,dark	0.82	0.92

Table2 Average Number of Associated Concepts and Time of association of the first, second and third concepts

Task	Number of Concept	Time of association(sec)		
		First	Second	Third
Higher	1.07	68	95	106
Lower	1.69	41	86	108
Synonim	0.78	40	67	86
Action	1.61	20	32	48
Enviro.	1.60	80	113	153

5 Concept dictionary construction using association experiment and its usage for information retrieval

We need a metric space for concept meaning to find whether the expression is metaphorical, literal or anonymous. For example, in T is V sentence, if T is located far from V in the space, the sentence has a big probability to be metaphorical or anonymous.

On the other hand, contextual analysis or disambiguation of plysemy are still difficult problems to be solved in natural language processing research. We have had, however, no good concept dictionary useful for such difficult probelms.

In this section, we conctructed a concept dictionary based on an association experiment, where 20 concepts are shown to 20 subjects. The concepts are 'sea','apple','traffic jam','slave', 'magician','family', 'antecedent','dictionary','network', 'knowledge','bog','egg', 'gold vein','mirror','hacker','close friend','adult', 'information', 'multimedia' and 'display'.

The tasks given to the subjects were to associate heigher-level concepts, lower-level concepts, synonims, features, action concepts related to the given concept, and environmental concepts related to the given concept. Average number of the associated concepts are shown in the Table2, where the average times of the association are also shown. Using T:the average time, j:order of the association, n:number of the subjects who associated the concept, and N:deviding n with the total subject number, cost C is formulated as follows;

$C = T \cdot a + j \cdot b + c/N$

where a, b, and c are weight coefficients. An example for associated action concepts are shwon in Table3.

Table 3 Cost of Action Concepts for Cherry Blossom
a=0.1,b=0.3,c=0.5

Concept	Cost
Fall	1.14
Bloom	1.27
See	1.61
Cultivate	5.75
Wave	5.90

Information retrieval system KMA(Keyword in Multi Ambience) was constructed which

used the concept dictionary to retrieve requested information from a Japanese dictionary. It can retrieve target information by a concept filtering method using features and related concepts in the concept dictionary even if the target is not defined precisely. KMA has an interface of GUI on X11 so that users can retrieve information interactively and flexiblly.

References

[1] Ishizaki,S:National Language Processing,Shokodo,1995

[2] Iwata,T. et.al:Discussion on the role of noun located at the object position of verb,Proc. of the 12th Meeting of JCSS, PA-8, pp.62-63,1995

[3] Ishizaki,S and Imai,M:Cognitive process of metaphor and its computer system, in Meterpher Research Group ed, "Creation and reframing of meaning through metaphor," Keio University,1995

[4] Tanaka,S:Conversation as a chain of word fragments, Gendai-eigo- kyoiku, No.8,pp.45-47, 1995

[5] Imai, M: Asymmetry in the taxonomic assumption: word learning vs. property induction.Child Language Research Forum, 27. Stanford University Press, 1995

[6] Imai,Y and Ishizaki,S:Discussions on semantic structure description of concent dictionary for metaphor understanding , Proc.of 51th Meeting of IPSJ,7H-2, pp.3-127-128,1995

[7] Ohkuma,T and Ishizaki,S:Functions of flexible dictionary retrival,Proc. of 51th Meeting of IPSJ, 5R-5, pp.3-113-114, 1995

[8] Ishizaki,S,Tanaka,S. and Imai,M.:Introducing metric space in concept dictionary for flexible semantic analysis, Natural Language Processing Research Group, JCSS, Vol.100, No.3,1994

[9] Tanaka,S:Cognitive science study(2)–Range of metaphor, Language-Culture- Communication(Keio University), 13, pp.69-111, 1994

[10] Tanaka,S. and Ishizaki,S:Sence generation at everyday conversation— role of 'you know' and 'I mean', Natural Language Processing Research Group, JCSS, Vol.100, No.3, 1994

[11] Imai, M., Gentner, D. & Uchida, N: Children's theories of word meanings: The role ofshape similarity in early acquisition, Cognitive Development, 9. pp.45-75, 1994

[12] Imai,M:Problems in lexical learning by foreign language learner—from the point of semantic phenomenon—, Study of Educational Psychology, Vol.41, pp.245-253, 1993

Keywords : cognitive semantics, metaphor understanding, concept dictionary, sense-making, polysemy, semantic component