Analysis of paraphrase and cohesion in seminar dialogues

Kikuko NISHINA, Yooko SASAGAWA, Mitsuru DOI
International Student Center
Tokyo Institute of Technology
2-12-1 Ookayama Meguro-ku Tokyo 152,Japan

1. Introduction

We approached structures of seminar dialogues to analyze special features of dialogues for computer use. We have made 20 video recordings (20 hours) of post graduate seminars in the Engineering Department at the Tokyo Institute of Technology.Through observation of the video recordings, we analyzed dialogues between foreign students and Japanese professors, focusing on cohesion and coherence in the seminar dialogues. Through this analysis, we have noticed that we are required to use some different methods of analysis for spoken dialogues to those for written discourses.

We propose to use lexical collocation to identify a word in a spoken discourse. For that purpose,we examined lexical relationships between words by way of a concept-based dictionary authored by EDR(the Japan Electric Dictionary Research Institute). We also considered utterance functions in seminar dialogues by defining and marking discourse markers.

2. Special features of seminar dialogues

We define seminar dialogue as a mutual understanding process to attain new information or knowledge, and differentiate it from idle chatter or strategic dialogue in sales talks.

We further define cohesion as word relationships within the text and coherence as word relationships with reference to background knowledge (the world).

Fig 1 shows the structure of dialogue divided into two frames. The first frame is one of knowledge that shares background knowledge or coherence. The second frame is one that shares the present dialogue content or cohesion. Cohesion binds lexical expressions in interaction between two speakers. On the other hand, coherence binds logical relationships including background knowledge of each speaker. Participants of a dialogue can reach mutual understanding by using both linguistic knowledge (or lexical rhetoric) and background knowledge.

However,it frequently happens that one particular participant has more knowledge or linguistic knowledge than the other as is the case with Japanese Professors and foreign students in this data. We found the phenomenon occurred for the following reasons, through observing dialogues between foreign students and Japanese professors or other staff.

(1) physical reasons of both speakers and listeners (physical disability : difficulty in hearing or speaking)

(2) language abilities of both speakers and listeners(children vs. adults,foreigners vs. native speakers)

(3) background knowledge of both speakers and listeners(experts vs. non-experts)

(4) inferential power for general matters of both speakers and listeners

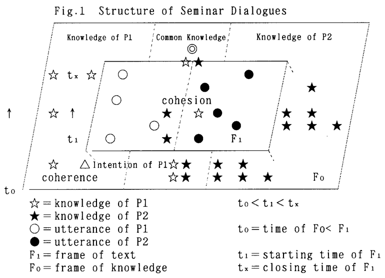

Figure 1:

A speaker in a dialogue often uses paraphrases as a rhetorical strategy, when he wants to persuade the audience. He also uses them to repair a communication breakdown when he wants to make them understand.

3. Rhetoric of dialogue strategies

In order to analyze spoken text, we have to treat it in proper way. We cannot use tools such as morphological or syntactic analysis for written texts because spoken sentences are not uttered in the same way as in texts. It can be said that such sentence are ill-formed or imperfect, but they occur frequently in spoken discourse. We thus have to consider special analysis methods for spoken discourse. Next, we turned our attention to lexical structures in spoken discourse. On the one hand, written sentences are usually formulated by well-formed formulas. On the other hand,spoken utterances do not conform to a written grammar, in that we often do not complete a sentence, or use a sentence out of its grammatical order. We decided to observe collocation of words in spoken text focusing on paraphrased parts of the dialogues. We aimed to find cohesion in the spoken discourse by observing lexical collocations, and consequently noticed the following patterns of cohesion within paraphrases in dialogues of the data.

(a)substitutions (b)ellipsis (c)reference (d)synonymy (e)hyponymy

(f)metonymy (g)syneodoche (h)antonymy

Those words paraphrased should be close to others in the sense of conceptual relationships. On this premise, we made use the concept-based dictionary produced by EDR. We measured the distance between words and the paraphrased words. We made conceptual network maps based on dialogue data. Fig2 shows such a conceptual network map based on part of a dialogue between a foreign PhD student and a Japanese professor, talking about signal processing.

We set concept level 0 to be the total concept, and placed it at the top of the map, and put words as terminals. These can be seen in the square frames. Some levels of the conceptuality are put between this top node and the terminals. We think that a definition of an undefined word in an utterance will be easier by referring to words that are uttered within close distance in the map, and focusing on statistic calculation.

student 1: chotto zutsu ,anoo ,tatoeba low pass filter dattara,chotto zutsu ,sono, smoothing ga okonawarete itte.

professor 2: un.

Fig.2 a Concept map of a dialogues

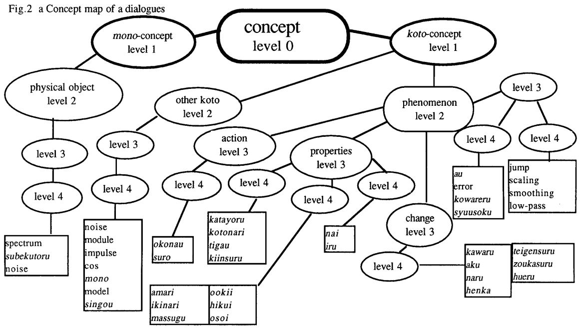

Figure 2:

student 3: ikinari shingoo kawaru no ga moo, sono, model, model tteiuka,bubun ga nai wake desu yo.

professor 4: un.

student 5: hijyoo ni shuusoku ga osoino de, sono noise no hikusa nimo kiin shiteiru n desu yo.

For example, "low pass filter","smoothing","shingoo"(signal)","shuusoku suru" and "noise" are concepts relating to signal processing ."Noise ga hikui."(noise is low.) and "shuusoku ga osoi."(convergence is slow.) are utterances that composed clauses. One can guess the word "hikui" and what the speaker has tried to say, even if the complete sentence cannot be distinguished clearly ,because the concept of this matter is known. It should be possible to refer to a word by calculating its distance to closely related concept words, by way of computer.

4. Functional Utterances as discourse markers

On the other hand, we observe the structure of coherence by dividing two types of functional words as discourse markers. One type is divided into additives, adversaries, casual, temporal, exemplification and summing up. We note functional words such as conjugations and demonstratives in the context as meta markers of coherence. Another type of functional words are those which indicate the speaker's intention. We identified and categorized these functions as receiving signal markers, requiring information markers, presenting information markers and so on.

5. Conclusion:

From this analysis, we conclude the following.

(1) In analyzing spoken discourse, we have to use different methods from hose used for written text. It seems a useful method to measure collocations of cohesion in the text to use a concept-based dictionary, although we need more data to accurately measure the conceptual distance between words within texts.

(2) We can observe spoken discourse from different viewpoints such as cohesion and coherence. Cohesion shows collocations of words or concepts in the discourse. Coherence shows its logical structure.

The following topics form the basis of extended further research:

(1) We will collect more data for analysis, to measure conceptual distance with a concept-based dictionary.

(2) We will form a descriptive rhetorical structure model according to conceptual relationships using algorithms.

(3) We will form protocols of discourse structure at seminars (comparing them with idle chat, and strategic talks)

Published Papers

[1] Kikuko NISHINA,Yooko SASAGAWA and Mitsuru Doi (1994)Analysis of Dialogues at Seminars - Towards Building a Syllabus for Technical Communication -Nihongo Kyooiku(Japanese language Teaching)84

[2] Kikuko NISHINA(1994,1995)Evaluation of EDR Concept Dictionary -from the language learners point of view-

Keywords: coherence, cohesion, collocation, background knowledge, concept-based dictionary, discourse markers