Linguistic Modeling of Dialogue Management

Yukinori TAKUBO, Satoshi KINSUI and Setuko ARITA
Faculty of Letters, Kyushu University
6-19-1 Higashiku,Fukuoka, 812
e-mail:takubltm@mbox.nc.kyushu-u.ac.jp

1. Introduction

We attempt to construct a performance model of speech, modeling the nature of operations the speaker/hearer performs in speech recognition and interpretation, with the limitation of memory and various distracting factors impeding ideal delivery and perception. For this purpose, we have developed a theory of discourse management, a version of Mental Space Theory, modified to accommodate dialogic discourse by incorporating a memory management system.

We posit a cognitive interface between linguistic expression and knowledge-base. This interface contains pointers or indices linked to addresses in the knowledge base, controlling access paths to the data in the base. It also functions as a working memory buffer for the production and interpretation of utterances. Thus the interface controls the input-output operations in speech. Utterances in a dialogue exchange can be redefined as input-output operations via this interface: registering, searching, editing, etc.

We hypothesize that there are overt reflections of these mental operations in the forms of discourse particles such as interjections and fillers. We assume that they work as mental monitoring devices for these discourse management operations, thereby supporting the speaker in the speech performance within the limit of memory resources and various distracting factors.

We have collected natural spoken data and examined the nature of discourse management operations by analyzing the uses of various discourse particles and interjections, within the model proposed in this study.

2. The proposal

The main theses of our approach to discourse management are as follows:

1) Speech can be divided into two parts, the part that codify the data to be communicated, and the part that codifies the mental operations to be performed for the data to be processed.

2) The operations coded in the various forms are to be defined as performing input-output operations on the database of the speaker and not that of the hearer's model in the speaker. We contend that the hearer's model in the speaker is not only unnecessary but also harmful in the description sentence forms.

3) We divide the interface into two components, I-domain and D-domain. The former is linked to temporary memory, houses the assumptions and propositions newly introduced to the discourse yet to be incorporated into the database and can be accessed only indirectly by inferences, logical reasoning, hearsay, and data search. The latter

is linked to the permanent memory, houses information already incorporated in the database and can be directly accessible by simple memory search like pointing to an index. New information passes through only via I-domain.

4) The real time processing of utterances can be modeled as interpreting and performing the operations in the linguistics forms that belong to the operative component in the sense of (1) and defined in terms of I-domain and D-domain.

In order to illustrate our proposals, we have chosen four topics from the operative component to analyze the nature of the operations involved. The topics are, vocatives, interjections, sentence final particles, and topic markers.

The forms can be illustrated in the following sentences.

Nee, Ichiro-kun, kimi-wa Yamada-kun to tomodati datta yo-ne.
Interjection you-TOP Mr.Yamada with friend be-Past SFN
(Say, Ichiro, I understand you are a friend of Yamada, right?)

'Nee' is an interjection serving as an attention getter, 'Ichiro-kun' is a vocative also working here as attention getter, 'kimi-wa' is a topic phrase, and 'yo-ne' at the end of the sentence is a concatenation of two sentence final particles. The message, or the data part of the sentence to be communicated is, 'kimi-ga yamada-kun to tomodati-de-aru koto'(the fact that you are friends with Yamada). All the rest of the sentence works as some sort of instructional cues as to how the sentence should be interpreted and incorporated into the data base. Let us see how this can be achieved.

3. Interjections

'Nee' is an attention getter used to ask for the attention of the hearer but the choice 'nee' as opposed to other such interjections such as 'ano, eeto, de' serves to anticipate the content of the utterance. 'Ne' in this case indicates that the information is new, and not the continuation of the previous utterance, which function is shared in 'ano' and 'eeto' but not when 'de' is used instead. While 'ano' and 'eeto' indicate that the speaker is in the process of building the content or editing the forms of the utterance, 'ne' shows that the speaker is ready to impart the content to the hearer. The difference between 'ano' and 'eeto' is more subtle, the former is used when the speaker is trying to extract linguistic information from the memory base, such as the name of a content or how the content can be properly coded to meet the needs of a particular situation, and the latter is used when the speaker is trying to secure a working buffer space in order to perform mental operations, such as calculation, memory search, content building, etc. It is our contention that interjections are the reflections of the various mental operations performed by the speaker as s/he utters a sentence and that, they, therefore can serve as cues for the hearer to guess and anticipate both the contents and the forms of the ensuing utterance.

4. Sentence-final particles

The sentence final particles can be interpreted as instructions as to how the content can be processed. We have chosen 'ne' and 'yo' to illustrate the nature of the operations involved. These particles have so far been treated as involving the speaker/hearer information exchange. 'ne' has been characterized as indicating that the information is known to the hearer, while 'yo' has been characterized as showing that the information is not new to the

hearer. We have demonstrated that this conception of the two particles cannot be true. Firstly, the introduction of the hearer's knowledge domain in the definitions of particles involves unnecessary complications leading to infinite regressions related to infinite embedding of shared knowledge. Secondly, these particles have usages that have no reference to hearer's knowledge. In the case of 'ne', it can be used to indicate self confirmation, self doubt, etc. 'yo' also can be used to show expression of one's emotion without necessarily presupposing the existence o f hearers. Thirdly, the concatenated form 'yo-ne' can never be explained by the existence and absence of hearer's knowledge. Thus, it is desirable both theoretically and empirically that the two particles be defined without reference to the hearer's domain of knowledge. We have characterized 'yo' as indicating the introduction of a proposition in the I-domain and 'ne' as a marker for indicating that the speaker is in the process of computation for verification. The proposition introduced in the I-domain ignites inferences of various kinds. Thus 'yo-ne' takes the proposition to be verified and serves to indicate the speaker is in the process of verifying the proposition by reference to some evidence available. Our approach views dialogue as a sort of verification game, a game aiming for maximal increase of verified knowledge on the part of the speaker. We have demonstrated that unlike other approaches that view speech as communication game involving maximal increase in shared knowledge, our approach is exempt from the danger of falling into the infinite regression and empirically more explanatory than other approaches which includes reference to hearer's domain.

5. Vocatives and topic markers

We have taken up two other topics, vocatives and topic markers. They are both shown to be markers for domain setting. Vocatives are to be characterized as setting the hearer domain for a particular dialogue session. We have hypothesized that the reference to the hearer's knowledge domain can be separated from the definition of linguistic form. The interface of the hearer's domain to the rest of the domain can be achieved by the setting up of the hearer's domain as a higher domain of discourse imposed onto the domain defined without recourse to the hearer's domain. The functions of vocatives are: to activate the relevant subpart of D-domain, to set up a channel between the speaker and the hearer and to reconfirm that the channel is still connected, and finally to set up the address terms used during the discourse. In Japanese, where the address terms can serve as nouns to refer to the second person, the terms are defined by vocatives. Thus a name like 'Ichiro-kun' or a family name like 'oto-san(father)' can serve as address terms and a noun directly referring to the hearer during a particular dialogue session if the person referred to by these nouns happens to be the hearer.

We have examined the marker 'wa' and have shown that it can also be characterized as indicating a mental operation. We have argued that the particle indicates the domain for predication. The main function of 'wa' is to separate an item in the database into two parts, the index part where the item to be predicated is identified with a minimum of information and the attribute part which is to be attributed to the item with that index. We have shown that this way of characterizing 'wa' can be extended to the other uses of 'wa' than topic, such as contrast and repetition, etc. and that it can solve mysteries of multiple usages associated with the particle.

6. Conclusion

We have developed a linguistic model of dialogue by modifying a mental space model

into dialogic discourse. The model seeks to model the real time processing of speech by humans by attributing procedural semantics to each linguisitic forms, thereby reflecting the processing of the speaker/hearer as they hear the words in the utterance in the sequence they appear in.

Keywords: discourse management, interjection, sentence-final particles, vocatives, topic marker