Workshop on Conversational Applications, June 2010,
Position Paper from NICT, Spoken Language Communication Group

Use Cases and Requirements for New Models of Human Language to Support Mobile Conversational Systems

(The position of the NICT, Spoken Language Communication Group)

Chiori Hori and Teruhisa Misu
{chiori.hori, teruhisa.misu}@nict.go.jp
Spoken Language Communication Group
National Institute of Information and Communications Technology (NICT), Japan

1. Participant profile

Our group is working on spoken language communication. Currently, the MASTAR project is ongoing in our group as follows:

MASTAR Project:

The MASTAR Project is a global hub for the speech and text research, with researchers in the field of language processing, machine translation and speech processing from Japan and abroad. It pursues research and development spanning the areas of multilingual speech-to-speech translation, multilingual text translation, multilingual dialog system technologies, and multilingual language resources, within a framework of collaboration between industry, academia and government. The MASTAR Project's goal is to collect and accumulate speech and text resources, a task that has proved difficult for individual organizations. The MASTAR Project will also accelerate the dissemination of research results back into society for practical application.

MASTAR Project Aims:

(1) to establish a global hub for speech/text resources and technologies.
(2) to initiate a sustainable system that stores and develops speech and text resources linking industry and society.
(3) Specifically…
   A. to carry out research and development and field tests, and to disseminate network-based speech-to-speech translation, one part of the Social Benefit Acceleration Projects of CSTP, Council for Science and Technology Policy.

B. to provide WEB 2.0 machine translation services for areas such as industry sectors and manuals, and to launch a positive development cycle for the accumulation of sharable dictionaries and corpuses for researching translation technologies.

C. to carry out research and development, field tests and dissemination of spoken dialog interface technologies in terms of universal communication to provide information for all users.

D. to create and disseminate global language resources.

(4) to set up an open R&D structure to promote research in collaboration with industry, academia and government, and to develop the human resources to support information technologies in Japan.

2. Use case of Speech Recognition and Spoken Language Understanding in the Kyoto Tour Guide Systems

We propose a spoken dialog system that adopt

- a statistic language model (N-gram) for speech recognition and
- a weighted finite-state transducer (WFST) for Spoken language understanding and management of dialog state.

The advantages of WFST-based description are:

- WFSTs can combine various statistical models for dialog management (DM), user input understanding, system action generation, etc.
- The system can select the best system action in response to a user input based on all combined models, where it can also accept multiple input hypotheses with confidence score to deal with uncertainty of the input.

Figure 1 shows an example of WFST-based Dialog systems [1].

Table 1 and Figure 2 show examples of XML and WFST Spoken Language Understanding (SLU) [2]. The XML defines user concept represented by keywords and key phrases are converted into a SLU WFST. SLU WFSTs can be composed with Scenario WFSTs representing context of dialog to understand user concept restricted by dialog states. Additionally, statistical models for concept understanding such as N-gram of dialog states can be applied to SLU WFSTs.

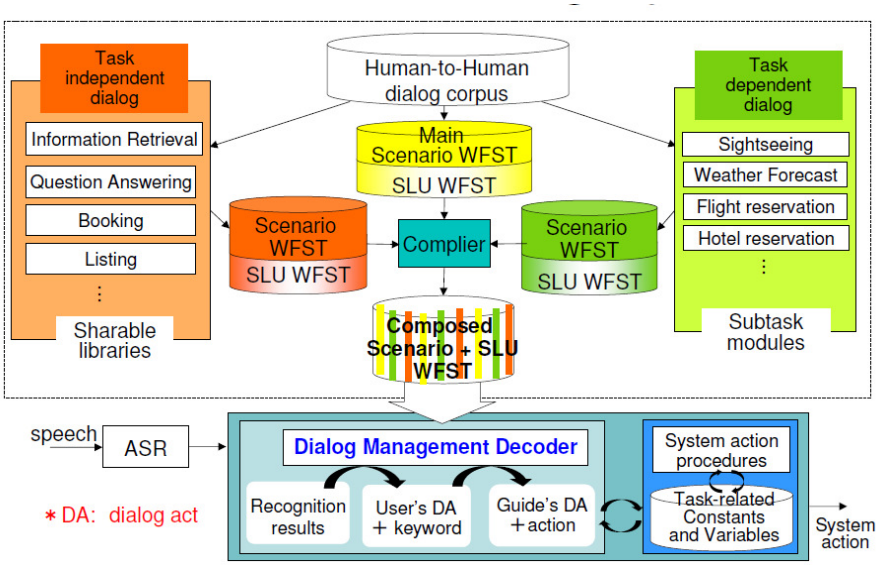

Fig. 1. WFST-based Dialog System

Table 1. An example of XML for SLU

<word-class label="station"> Tokyo Kyoto </word-class> <keyword-class label="origin"> (station) </keyword-class>	<keyword-class label="time"> six seven eight nine ten eleven twelve </keyword-class> <keyword-class label="destination"> (station) </keyword-class>	<plan repeat="true"> from,(origin) to,(destination) </plan> <depart> at,(time) </depart

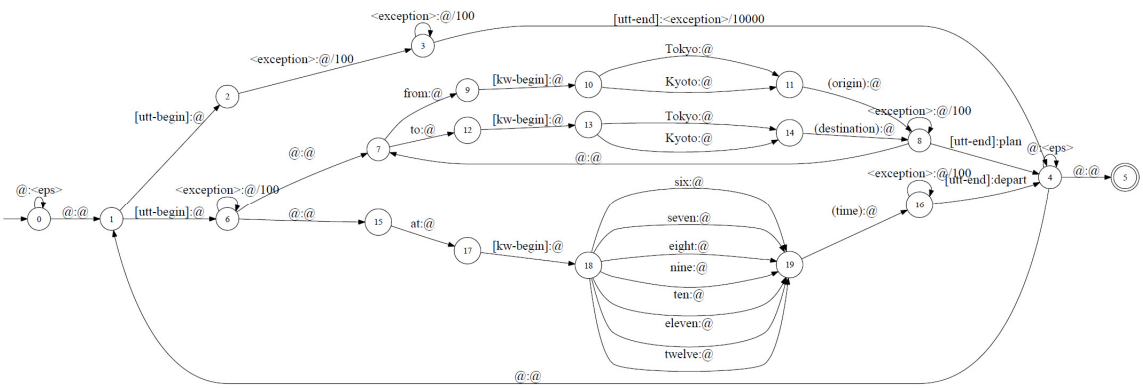

Fig. 2 An example of SLU WFST

3. Problems in implementing the above framework using SRGS/SISR

3.1. Context sensitive speech recognition

We would like to propose a framework of dialog context-based ASR. Statistical Language models for ASR are required to be tuned depending on the current dialogue context determined by previous system prompt, dialogue situations.

3.2. Separation of Speech recognition and Natural Language Understanding

We need to implement speech recognition systems which are more robust to natural language expressions. N-gram language models can be a solution. Consequently, we will need a framework to label semantic annotations on ASR results, afterward.

3.3. Spoken Language Understanding using WFSTs

To realize context sensitive semantic annotation for SLU, we need a description for WFST.

Reference:

[1] Chiori Hori, Kiyonori Ohtake, Teruhisa Misu, Hideki Kashioka, Satoshi Nakamura, "Statistical dialog management applied to WFST-based dialog systems," ASRU2009

[2] Chiori Hori, Kiyonori Ohtake, Teruhisa Misu, Hideki Kashioka, Satoshi Nakamura, "Recent Advances in WFST-based Dialog System," Interspeech 2009.