Starting with this issue, we will be publishing our annual summary of research on July 1 of each year to coincide with MIT’s fiscal year calendar. Therefore, this transitional issue covers the eighteen months period from January 1997 to June 1998.

**The Spoken Language Systems Group**
The Spoken Language Systems (SLS) Group was established at the MIT Laboratory for Computer Science in February 1989. For readers not familiar with the goals of our research, motivations of our approaches, strategies that we have adopted, and technology and systems that we have developed, I refer you to our web page (http://www.sls.lcs.mit.edu) for background information.

Briefly stated, the group’s goal is to develop the underlying human language technologies that will enable universal access of information, anytime, anywhere, by anybody. We believe this capability calls for a new generation of speech-based interfaces that enable the computer to serve as a conversational partner to the user.

Figure 1 shows the major components of a typical conversational system. The spoken input is first processed through the speech recognition component. The natural language component, working in concert with the recognizer, produces a meaning representation. For information retrieval applications illustrated in this figure, the meaning representation can be used to retrieve the appropriate information in the form of text, tables and graphics. If the information in the input utterance is insufficient, the system may choose to query the user for clarification. Speech output can be obtained by processing the information through natural language generation and text-to-speech synthesis. Throughout the process, discourse information is maintained and fed back to the speech recognition and language understanding components.

---

**Figure 1.** A generic block diagram of a typical MIT conversational system.
Recent Highlights
During the eighteen months period, we continued our research and development in human language technologies and conversational systems on several fronts. We continued to improve the performance of our SUMMIT recognition system, moving beyond the anti-phone modelling to include near-miss modelling and segmentation by recognition (cf. Jane). This has resulted in our achieving state-of-the-art phonetic recognition performance on the TIMIT corpus. We continue to develop ANGIE, a hierarchical, morpho-phonological framework for representing speech and language, incorporating several sub-word layers for speech processing. ANGIE’s utility has been demonstrated in several areas, including letter/sound conversion (cf. Aarati), speech recognition and word spotting (cf. Ray), and durational modelling (cf. Grace).

With regard to system development, two aspects are particularly noteworthy. First, a new conversational interface called JUPITER came into being in 1997. JUPITER is a telephone-only system for weather information for more than 500 cities worldwide. The weather information is obtained from four on-line sources on the Web, and is updated several times daily. JUPITER employs Galaxy’s client-server architecture, except the client is simply a telephone. It serves as a platform for investigating several research topics. First, by using the telephone as a means of accessing the information, we can empower a much larger population to access the wide range of information that is becoming available. In the scenario that we envision, a user could conduct “virtual browsing” in the information space without ever having to point or click. Second, displayless information access poses new challenges to conversational interfaces. If the information can only be conveyed verbally, the system must rely on the dialogue component to reduce the information to a digestible amount, the language generation component to express the information succinctly, and the test-to-speech component to generate highly natural and intelligible speech. Third, channel distortions place heavy demands on the system to achieve robust speech recognition and understanding. Finally, by applying human language technologies to understanding the “content,” in this case the weather forecast, we can manipulate and deliver exactly the information that the user wants, no more and no less. Since May 1997, we have installed a toll free telephone number, so that naive users can call JUPITER and inquire about the weather. Log-files are created automatically, and the sentences are also transcribed as they come in. Even without active advertising, we were able to collect over 60,000 sentences from nearly 12,000 calls. The data have proved to be invaluable for system development, training, and evaluation. Following the success of JUPITER, we have started to develop other, similar applications, including flight status, local traffic, and urban navigation.

On another front, we have started a major redesign of our GALAXY architecture. In 1994, we introduced GALAXY, a client-server architecture for accessing on-line information using spoken dialogue. Since then, GALAXY has served as the testbed for our research and development of human language technologies, resulting in systems
in different domains and languages, and with different access mechanisms. In 1996, we made our first significant architectural redesign to permit universal access via any web browser. The resulting WebGALAXY system makes use of a “hub” to mediate between a Java GUI client and various compute and domain servers. Since GALAXY has recently been designated as the first reference architecture of the newly launched DARPA Communicator initiative, we are developing a new version of the GALAXY architecture, this time with the specific goals of promoting resource sharing and plug-and-play across multiple sites. To enable multiple system developers to experiment with different domains, components, and control strategies, we have decided to make the hub “programmable,” i.e., a scripting language controls the flow through each dialogue without recompiling. The hub communicates with the various servers via a standardized frame-based protocol. At this point, a new GALAXY-II architecture has been designed and implemented. In the coming months, we expect to make available to the research community the programmable hub, its associated libraries, and a set of human language servers, so that others can experiment with the system for technology and application development.

Last but not least, a record number of students received their degrees in this period. The research topics that they addressed include acoustic modelling, dialogue modelling, speaker adaptation, word spotting, speech synthesis, and tools development. A brief description of their work can be found in this document. Copies of their theses are available for download from our web page. One of our students (Jane Chang) received the Sprowls doctoral thesis award in computer science, while another (Kenney Ng) received the best poster award at the 1997 ACM SIGIR meeting and the ELSNET-ELRA Olympics best student paper award at Eurospeech 1997.