GALAXY: A HUMAN-LANGUAGE INTERFACE TO ON-LINE TRAVEL INFORMATION [1]

David Goddeau, Eric Brill [2] James Glass, Christine Pao
Michael Phillips, Joseph Polifroni, Stephanie Seneff, and Victor Zue

Spoken Language Systems Group
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139 U.S.A.

ABSTRACT

This paper introduces GALAXY, a distributed system for providing access and intelligent mediation for on-line information and services via human-language technology. We have chosen to focus our initial implementation on the travel domain (including air travel, local navigation, and weather) largely to leverage off our past experience in the development of the VOYAGER [1], ATIS [2], and PEGASUS [3] systems. GALAXY is designed to access real databases that are distributed among various information networks (e.g., Internet, LAN, or commercial networks such as Compuserve). Our initial system utilizes American Airlines' EAASY SABRE, the NYNEX Electronic Yellow Pages, the US Census Bureau's map database, and the National Weather Service database. GALAXY was developed as an environment for exploring important research issues in human language technology. Moreover, GALAXY will provide a demonstration of an interface that will enable users to efficiently access, process, and manipulate a vast amount of on-line information.

INTRODUCTION

The last decade has witnessed significant progress in Human Language Technology (HLT). For speech recognition (SR), word error rates continue to decrease, while barriers to speaker independence, continuous speech, and large vocabularies begin to fall. Similarly, new generation language understanding systems have achieved remarkable levels of sentence understanding in restricted domains. Moreover, a new breed of system has emerged that combines the two technologies to achieve speech understanding capabilities. Over the past five years, our group has been an active participant in the development of such conversational systems, i.e., systems that can understand and respond verbally in order to help a user access information and solve specific problems [1, 2, 3].

Such rapid technological progress raises several timely questions. When will this technology be available for productive use? What technological barriers still exist that will prevent large-scale HLT deployment? We believe that an effective strategy for answering these questions is to develop the underlying technologies within *real* applications, since this will force us to confront some of the critical technical issues that may otherwise elude our attention. Consider, for example, the task of accessing information in the Yellow Pages of a medium-sized metropolitan area such as Boston – a task that can be viewed as a logical extension of the VOYAGER system we developed in 1989. The vocabulary size of such a task could easily exceed 100,000, considering the names of the establishments, street and city names, and listing headings. A task involving such a huge vocabulary presents us with a set of new technical challenges. Among them are:

- How can adequate acoustic and language models be determined when there is little hope of obtaining a sufficient amount of domain-specific data for training?
- What search strategy would be appropriate for very large vocabulary tasks? How can natural language constraints be utilized to reduce the search space while providing adequate coverage?
- How can the application be adapted and/or customized to the specific needs of a given user?
- How can the system be efficiently ported to a different task in the same domain (e.g., changing the geographical area from Boston to Washington DC), or to an entirely different domain (e.g., library information access)?

There are many other research issues that will surface when one is confronted with the need to make human language technology truly useful for solving real problems. A particularly important issue concerns unknown words. In a real domain such as the Yellow Pages, a certain fraction of the words uttered by users will not be in the system's working vocabulary. This is unavoidable partly because it is not possible to anticipate all the words that all users are likely to use, and partly because the database usually changes with time (e.g., new restaurants opening up). Truly useful systems must be able to cope with unknown words simply because they will always be present, and ignoring them will not satisfy the user's needs. The system must be able not only to detect new words, taking into account acoustic, phonological, and linguistic evidence, but also to acquire them, both in terms of their orthography and linguistic properties.

Yet another example of important research issues illuminated by real applications is the use of dialogues in human-computer interactions. Human verbal communication is a two-way process involving multiple, active participants. Mutual understanding is achieved through direct and indirect speech acts, turn taking, clarification, and pragmatic consid-

[1] This research was supported by ARPA under Contract N00014-89-J-1332, monitored through the Office of Naval Research.

[2] Currently at Johns Hopkins University.

erations. Our experience with the development of the PEGASUS system for on-line travel planning convinced us that an effective spoken language interface for information retrieval and interactive transactions must incorporate extensive and complex dialogue modelling – initiating appropriate clarification subdialogues based on partial understanding, and taking an active role in directing the conversation towards a valid conclusion.

Finally, working on real applications offers a set of potential side benefits. While years may pass before we can develop unconstrained spoken language systems, we are fast approaching a time when systems with limited capabilities can help users interact with computers with greater ease and efficiency. We believe that the time is ripe for us to demonstrate the usefulness of the technology. Working on real applications thus has the potential benefit of shortening the interval between technology demonstration and its ultimate use. Besides, applications that can help people solve problems *will* be used by real users, thus providing us with a rich and continuing source of useful data.

This paper introduces GALAXY, a distributed system for providing access and intelligent mediation for on-line information and services via human-language technology. We have chosen to focus our initial implementation on the travel domain (including air travel, local navigation, and weather). GALAXY was developed as an environment for exploring important research issues in human language technology. Moreover, GALAXY will provide a demonstration of an interface that will ultimately enable users to efficiently access, process, and manipulate a vast amount of on-line information.

ARCHITECTURE

The design of GALAXY is based on the observation that the number of on-line information sources and services available is rapidly increasing. They typically reside at various points along the information highway (e.g., Internet, local area network, and commercial networks such as Compuserve), requiring different protocols for information access. At the same time, the number of people attempting to make use of these services is also increasing. As a result, the system must be distributed in nature.

The architecture of GALAXY is based on the model of a number of domain servers providing information for a collection of user clients, as illustrated in Figure 1. The domain servers encapsulate some area of expertise, and are each capable of dealing with a certain set of queries. These servers contain general knowledge about the structure of their domain in addition to the capability of accessing specific databases. They serve to insulate the user from the location and format of the data sources.

The client program provides the interface to the user. It captures input from the user, including speech, typed input, and pointer events. It also presents the servers' responses using graphics, text, and synthetic speech. Note that the human language aspects can be implemented as a server as well, thus removing the need for computationally expensive clients.

The client and server processes interact via a domain-independent communication protocol. This protocol transmits the input queries to the servers in a language independent form, and allows for a variety of response modalities back to the client.

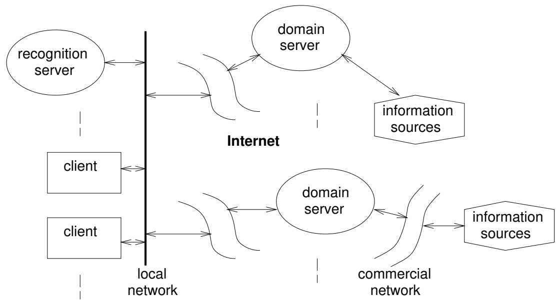

Figure 1: GALAXY Architecture.

The domain servers do the actual work of interpreting user requests, locating required information, and composing a suitable response. They correspond roughly to the backend and discourse management components of a conventional spoken language system. These are among the most difficult and time-consuming parts of a spoken-language system to produce, requiring a great deal of domain specific analysis and programming. One of the goals of GALAXY is to organize the system such that, once one of these knowledge domain servers is produced, it can be easily utilized by the widest community possible. This in turn will hopefully inspire other members of the community to produce other domain servers and make them accessible.

TRAVEL DOMAIN IMPLEMENTATION

We have chosen to implement an initial human language interface for GALAXY in the travel domain, building on our past experience in the VOYAGER, ATIS, and PEGASUS domains. It also has the capability of providing weather information. GALAXY makes use of real databases concerning airline schedules, Yellow Pages, metropolitan maps, and weather forecasts, and will help people make general travel plans.

Domain Servers

City Guide The City-Guide domain server provides tourist information and navigation assistance for a specific metropolitan area, in this case, the Boston area. Its current information sources consist of an on-line version of the NYNEX Yellow Pages for the 617 Area Code[3], an additional database of local points of interest, including a personal rolodex, and geographical information from a subset of the US Census Bureau's TIGER files[4].

The server can answer queries about general classes of objects ("Are there any Indian restaurants in Cambridge?"), provide information about a specific location ("What is the phone number of the Back Bay Hilton?"), and deal with a variety of constraints ("on Main Street", "near MIT," etc.). In addition to database access, it can perform data fusion and computation ("Give me directions from the Royal East to MIT," "How far is the nearest Chinese restaurant?", etc.).

[3] We have thus far only included headings that are related to tourism, such as hotels, restaurants, car rental agencies, theatres, etc. There are currently about 2000 listings.

Its response to the user includes a natural language reply, a mouse-sensitive list displaying the requested items, and a map display of relevant locations.

Airline Reservations The Airline Reservation server provides access to the American Airlines EAASY SABRE system. This is a consumer-oriented on-line reservation system available via modem from several commercial services. It currently has more than 700,000 subscribers.

The domain server connects to EAASY SABRE through its dialup interface, which was designed for human-machine interaction. The server must therefore analyze the returned screens and extract the important information for responding to the user's request. The implementation of this server was initially derived from the MIT PEGASUS system[3], although it currently lacks many of the more sophisticated features of that system. The current server allows queries about flight schedules, properties, availability and fares between 220 major cities throughout the world. Additional capabilities from PEGASUS will be included in future versions.

Weather Information The Weather Server provides access to several different sources of weather data originating with the US Weather Service, stored at a number of Internet sites and available via the Gopher protocol[9]. Currently, data is available for around 120 U.S. cities, and includes current forecasts and climatological data, temperature maps, and satellite images.

As well as supplying data directly to the client, the server can also process and analyze some of the text-based information sources, such as forecasts, to answer narrow queries ("What will the weather be like in Miami?", "Is it going to rain on Tuesday?", etc).

Speech Recognition

The recognition server uses the MIT real-time SUMMIT recognizer [5]. This recognizer uses a set of context-independent models, a class-bigram language model, a forward Viterbi search with fastmatch [6] and a backward A^* search to produce an N-best list of recognition hypotheses.

Lexicon construction was problematic because the usual source of the recognition lexicon, a large domain specific corpus, was not available. The lexicon for this implementation contains words derived from a number of language corpora including ATIS [7] and VOYAGER [8]. In addition, domain-specific vocabularies, consisting mostly of proper names, including cities and towns, street names, and business establishments, were extracted from the databases. This results in a vocabulary of over 4000 words.

Due to the lack of a training corpus and the multi-domain nature of GALAXY, we were unable to follow the conventional methodology of training the acoustic and language models. Instead, the acoustic models were trained from a collection of sentences from the ATIS and VOYAGER spontaneous speech corpora. The language model is a class bigram constructed and trained from subsets of the ATIS and VOYAGER corpora, along with database listings derived from the Yellow Pages database. The within-class probabilities were derived from the databases (i.e., Yellow Pages and EAASY SABRE) rather than from the corpora. This technique represents a first step toward a method of producing language models for new domains without collecting a new corpus, by extracting a "general" language model from existing corpora and customizing it using the databases of the new domain.

Natural Language Understanding

The natural language processing for GALAXY makes use of the TINA formalism [11]. In the current implementation the parser is compiled into the client program, since it imposes very little computational overhead. The parser processes the N-best list returned by the recognizer, selecting the first parsable hypothesis. If such a hypothesis is found, a semantic frame representing the meaning is derived from the resulting parse tree. The client must decide which server is appropriate for answering each user query, by examining relevant entries in the semantic frame.

The grammar for GALAXY was constructed by first combining existing grammars for PEGASUS and for VOYAGER, and augmenting them with a set of new rules specific to the weather subdomain. In designing these domain-specific grammars, we had tried to maintain a consistent set of top-level syntactic rules covering the variety of ways generic database-access queries can be expressed in English. The domain-specific concepts occur mainly in the low-level rules covering particular semantic classes of noun phrases, etc., within each subdomain. This strategy made it relatively easy to merge all the rules into a single rule set.

Once this initial rule set had been constructed, we were left with the problem of parsing the large number of proper names in GALAXY. These names could in principle be enumerated from the external database and specified explicitly in the rules. Many of them involve a mix of common and unusual words, such as "Popeye's Famous Fried Chicken," or "Bartley's Burger Cottage." Users would see these names in the display window, and could refer to them from the complete listing or only through partial reference, such as "Popeye's" or "Popeye's Fried Chicken." It was straightforward to write rules to cover *unknown* words in proper names, inferring the semantic class from appended generic words when available, as in "Asmara Restaurant." However, it was more difficult to devise a scheme for parsing names that included a mix of known (elsewhere in the rules) and unknown words. Ultimately, a successful strategy was developed for using any single unknown word as a "seed," licensing its neighbors to also be unknown, successively, until a successful parse was recovered. We verified this strategy by parsing "Show <Named Establishment>" for all the establishments within the Yellow Pages database available to GALAXY.

Response Generation

The language generation component of GALAXY, GENESIS, is used both to paraphrase an input query and to supply verbal responses to user requests. GENESIS is configured so that the language-dependent aspects of generated strings are controlled via external tables, with the procedures operating language-independently from a semantic frame representation common to all languages. A detailed description of GENESIS can be found in [10].

The paraphrase provides convenient confirmation that the system has properly understood the user query. Paraphrasing also provides the possibility of translating user queries among the various languages supported by the system. We have begun to develop such paraphrasing capabilities within GALAXY for Spanish, Mandarin Chinese, and Japanese.

Responses are often composed from a mixture of information derived from the semantic frame and from the database. For instance, a typical response might be: "There are fourteen Chinese restaurants in Cambridge listed." Generation currently resides with the servers rather than with the client, and each subdomain maintains separate generation tables.

CURRENT STATUS

At this writing, the initial implementation of GALAXY in the travel domain has been completed. The human language interface, including speech recognition and language understanding, runs in near real time on a high-end workstation. Overall system response time is constrained by database access time, which can depend on user load and network traffic.

Since we have not yet collected any data in this domain, we are presently unable to formally evaluate the system's performance. Nevertheless, we have trained and evaluated an initial version of a GALAXY recognizer on a combination of ATIS and VOYAGER corpora. Initial evaluations show word error rates of 19.3% and 13% for ATIS and VOYAGER development sets, respectively. These results show a moderate increase in word error rate over our domain-specific word error rates of 16.1% for ATIS and 12.6% for VOYAGER, even though the vocabulary size has increased significantly from the original ATIS and VOYAGER domains (approximately 2500 words forATIS and 380 for VOYAGER).

DISCUSSION AND FUTURE WORK

As we stated at the outset, the GALAXY project provides both a driving force and experimental testbed for exploring many important research issues in spoken language technology. Moreover, it potentially offers practical utility to a broad user community, helping them efficiently access, process, and manipulate a vast amount of on-line information. However, the system described in this paper is in its infancy, and much work remains to be done. Besides inproving the core language technology – speech recognition, language understanding, and language generation – we plan to address the following set of research issues.

Discourse and Dialogue Modeling We believe these are two crucial yet insufficiently explored technical areas. This issue is particularly important for "displayless," telephone-based transactions.

Dealing with Unknown Words Unknown words are inevitable in any real application. We must develop procedures for detecting and learning new words (both their phonological and linguistic structures), and rejecting out-of-domain queries.

Portability We will investigate procedures that will enable us to port to new domains from at most a small amount of domain-specific data. Of particular interest are the acquisition of a new lexicon, language model, grammar, dialogue model, and understanding capability. In addition, we will continue and expand our effort in developing multilingual conversational systems.

In the near future, we plan to develop the necessary infrastructure so that we may begin to collect data from naive users when they interact with GALAXY. The resulting data will allow us to refine the components of the system, particularly for discourse and dialogue modelling, and to evaluate system performance as improvements are introduced (such as context-dependent acoustic modelling and adaptive language modelling).

We also plan to continually expand the capabilities and features of GALAXY in the travel domain by adding new information sources (e.g., flight arrival information, Yellow Pages for another city) and new servers (e.g., providing navigational assitance from one city to another). Eventually, we would like to expand into other domains, such as library access and medical information retrieval. A major motivation behind broadening the scope is to drive our research on portability and domain-independence in our component technologies. As the system becomes more stable, we will try to make it available to real users and researchers. A user community would not only supply a continuous source of speech and language data for system development, but also allow us to gain insight into issues such as discourse and dialogue structure in human-computer interactions.

ACKNOWLEDGEMENTS

Part of the work described in this paper is conducted in collaboration with American Airlines SABRE Travel Information Network. We would like to thank NYNEX for providing us with a copy of their electronic Yellow Pages.

REFERENCES

[1] Glass, J., Goodine, D., Phillips, M., Sakai, S., Seneff, S., and Zue, V. "A Bilingual VOYAGER System," *Proc. Eurospeech*, pp. 2063–2066, September, 1993.

[2] Zue, V., Glass, J., Goddeau, D., Goodine, D., Hirschman, L., Phillips, M., Polifroni, J., and Seneff, S. "The MIT ATIS System: February 1992 Progress Report," *Proc. DARPA Speech and NL Workshop*, Harriman, NY, pp. 84–88, February, 1992.

[3] Zue, V., Glass, J., Goddeau, D., Goodine, D., Pao, C., Phillips, M., Polfroni, J., and Seneff, S. "PEGASUS: A Spoken Dialogue Interface for On-line Air Travel Planning," *Proc. International Symposium of Spoken Dialogue*, Tokyo, Japan, pp. 157–160, November, 1993.

[4] Bureau of the Census, *TIGER/Line^{TM} Files, 1992 Technical Documentation*, Washington DC: The Bureau, 1993.

[5] Zue, V., Glass, J., Goodine, D., Phillips, M., and Seneff, S., "The SUMMIT Speech Recognition System: Phonological Modelling and Lexical Access," *Proc.ICASSP-90*,Albuquerque, NM, pp. 49–52, April, 1990.

[6] Phillips, M. and Goddeau, D., "Lexical Tree Based Pruning for Segment-Based Speech Recognition", *These Proceedings*, September, 1994.

[7] MADCOW, "Multi-Site Data Collection for a Spoken Language Corpora", *Fifth DARPA Speech and Natural Language Workshop*, pp. 7–14, February, 1992.

[8] Zue, V., Daly, N., Glass, J., Goodine, D., Leung, H., Phillips, M., Polifroni, J., Seneff, S., and Soclof, M., "The Collection and Preliminary Analysis of a Spontaneous Speech Database", *DARPA Speech and Natural Language Workshop*, pp. 160–167, October, 1989.

[9] Alberti, B., Anklesaria, F., Lindner, P., McCahill, M., and Torrey, D., "The Internet Gopher Protocol: a distributed document search and retrieval protocol", University of Minn. Microcomputer and Workstation Networks Center, Spring 1992.

[10] Glass, J., Polifroni, J., and Seneff, S. "Multilingual Language Generation Across Multiple Domains," *These Proceedings*, September, 1994.

[11] Seneff, S. "TINA: A Natural Language System for Spoken Language Applications," *Computational Linguistics*, Vol. 18, No. 1, pp. 61–86, 1992.