

THE USE OF A COMMERCIAL NATURAL LANGUAGE INTERFACE IN THE ATIS TASK

Evelyne Tzoukermann
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

Abstract

A natural language interface for relational databases has been utilized at AT&T Bell Laboratories as the natural language component of the DARPA ATIS common task. This is a part of a larger project that consists of incorporating a natural language component into the Bell Laboratories Speech Recognizer.

The commercially available system used in this project was developed by Natural Language Incorporation (NLI), in particular by J. Ginsparg [Ginsparg, 1976]. We relate our experience in adapting the NLI interface to handle domain dependent ATIS queries. The results of this allowed the exploration of several important issues in speech and natural language:

1. the feasibility of using an off-the-shelf commercial product for a language understanding front end to a speech recognizer,
2. the constraints of using a general-purpose product for a specific task.

1 Introduction

The ATIS common task was designed by DARPA and the members of the DARPA community to build and evaluate a system capable of handling continuous and spontaneous speech recognition as well as natural language understanding. Although the evaluation task is still not fully defined, the ATIS common task presents the opportunity to develop reliable and measurable criteria. The present paper focuses on the natural language component only, the integration with speech being reported in other papers [Pieraccini and Levin, 1991]. The domain of the task is on the Air Travel Information Service (ATIS). The project touches on a wide range of issues both in natural language and speech recognition, including incorporation of an NL interface in speech understanding, flexibility in the type of input language (i.e.

spoken or written), relational databases, evaluation of system performance, possible limitations, and others.

The NLI system¹ is driven by a syntactic parser designed to handle English queries that are characteristic of the written language. In contrast, ATIS syntax is characteristic of spoken and spontaneous language. Therefore, one of the primary questions in using the NLI system has been how to overcome problems related to the discrepancy between written and spoken language input. Issues related to the ATIS domain and queries on the one hand, and to the construction of the NLI interface on the other hand are addressed. The task of the experiment is then described along with the results.

2 Why use a commercial product?

Using a commercial product is attractive for a number of reasons:

- within Bell Laboratories, there has been no effort so far to develop a natural language interface (although this may change). Therefore, it is a significant savings of time and effort to use a publicly available system in order to achieve the larger task, that is the integration of speech and natural language.
- within the task of language understanding, the use of a natural language interface meant to understand written language input, exposes issues specific to speech incorporation.

3 NLI system description

The NLI system is composed of a series of modules, including a spelling corrector, a parser, a semantic interface consulting a knowledge representation base, a con-

¹The acronym NLI should not be confused with the suffix of the transcription sentences “.nli”, meaning natural language input.

versation monitor, a deductive system, a database inference system as well as a database manager, and an English language generator [NLI Development Manual, 1990]. The components are:

- a **spelling corrector** which analyses morphological forms of the words and creates a word lattice.
- a **parser** which converts the query represented by a word lattice into a grammatical structure resembling a sentence diagram.
- a **semantic interface** which translates the parser output into the representation language, a hybrid of a semantic network and first-order predicate logic. This language permits representation of time dependencies, quantified statements, tense information, and general sets. The representation produced by this system component is concept-based rather than word-based.
- a **language generator** which transforms the representation language statements into an English sentence.
- an **interpreter** which reasons about the representation language statements and makes decisions. It is also called by the semantic interface to help resolve reference, score ambiguous utterances, and perform certain noun and verb transformations.
- a **database interface** which translates the representation language statements into SQL database statements.
- a **dictionary** which contains over 9,000 English words and their parts of speech.
- a **set of concepts** which consists of internal notions of predicates, named objects, and statements.
- a **set of rules** which consists of “facts” that make up the rule base of the system. Rules are statements which are believed to be true. The rule interface can handle quantification, sets, and general logic constructs.

4 Making the connections

The first steps in using NLI consisted of creating connections within the concept tables of the database and in reformatting the ATIS database into the NLI formalism. This has required different types of operations; one operation consisted of taking a relation, naming what it represents and connecting it with its properties. For

example, the relation “aircraft” represents plane, and attribute “weight” in aircraft represents the weight of the plane or how heavy or light it is with units of pounds. Another was to instantiate verb templates. For example, verbs such as “travel”, “fly”, “go”, etc. must be linked to a filler such as “from_airport” through the preposition “from”. The relation “flight” contains information about which airlines (relation “airline” via “airline_code”) fly flights (relation “flight”) from airports (“from_airport”) to airports (“to_airport”) on airplanes (relation “aircraft” via “aircraft_code”) at times (“departure_time”, “arrival_time”, “flight_day”). A third type of expansion involves the synonymy between two items; for example, the system must be informed that the string “transportation” should be understood as “ground transportation”. Connections were added incrementally in order to expand the coverage.

5 System performance and analysis of results

The system has been trained on the set of training sentences (only the non-ambiguous sentences called “class A”, i. e. about 550 sentences) recorded at Texas Instruments and the set of test sentences (i. e. about 100 sentences distributed by NIST) that were used for the June 1990 DARPA task. The last test made on the training sentences gave over 61% successfully answered queries which conformed to the Common Answer Specification (CAS) required by the NIST comparator program. It must be pointed out that the translation of NLI output answers into the CAS format was not a straightforward process. For example, when the system could not answer a query successfully, it output various expressions such as: *Sorry, I didn't understand that. Please check your spelling or phrasing*, or *The database contains no information about how expensive airports are*, or *I could not find a meaning for the noun “five”*, etc., so finding the correct CAS format became guess work. For this purpose, a program was written by Mark Beutnagel² at Bell Laboratories to handle the general cases (transformation of the output tables into the CAS tables) but also a number of idiosyncratic ones.

The February '91 test was designed to handle different types of queries, unlike the June '90 test that had only class A sentences. The queries were divided in four categories: class A or non-ambiguous, class AO non-ambiguous but containing some so-called verbal deletions (they have in fact all sorts of spoken-specific language peculiarities, such as *What flights list the flights*

²I want to thank Mark Beutnagel for his many hours of useful help.

from Pittsburgh to San Francisco?), class D for dialog-sentences where queries are presented by pairs (one member of the pair indicating the context of the sentence, the other member being the query itself), and class DO for dialog sentences with verbal deletions. At the time of the experiment, although NLI could handle queries with anaphoric pronouns across sentences such as in the pair *Show the flights from Atlanta to Baltimore. When do they leave?*, the connection file had not been shaped in that direction. The system was trained only to handle the class A queries. Answers to the four categories were run and sent, but only the class A results are of real interest and relevant. The following table shows the results of the four sets of sentences. The queries were evaluated in three different categories, "T" for "True", "F" for "False" and "NA" for "No_Answer":

	CLASS A	CLASS AO	CLASS D	CLASS DO
T	69	2	17	0
F	60	8	18	2
NA	16	1	3	0

6 Error analysis

The first obstacle encountered in utilizing NLI was the nature of the input queries. The ATIS task is meant to understand spoken and spontaneous language whereas NLI is built to understand written type of language. There are a number of discrepancies between spoken and written language that involve a different parsing strategy; spontaneous speech contains various kinds of:

- **repetitions** such as *through through* in the sentence *Please show me all the flights from DFW to Baltimore that go through through Atlanta;*
- **restarts** as shown in the query *What flights list the flights from Pittsburgh to San Francisco;*
- **deletions** such as the missing word *Francisco* in *Display ground transportation options from Oakland to San downtown San Francisco;*
- **interjections** with the use of the word *Okay* in *Okay I'd like to see a listing of all flights available...;*
- **ellipsis** such as in the third phrase *I'm sorry cancel that flight. The passenger wants to fly on Delta. How about Delta 870 on the 12th?;*

Note that in this format, the punctuation marks which might give the system information do not occur.

There are a number of explanations for the unanswered sentences:

- **Lexical gaps:** if a lexical item is not in the lexicon, no analysis is given. The problem in lexical gaps is partly due to the domain specific vocabulary of the ATIS task. In the following example *I need flight schedule information from Denver to Philadelphia*, the system does not have the word *schedule* in the lexicon; therefore the sentence is rejected.
- **The informal addition of information** is common to spoken language, more than written language. For example, in the following sentence, the speaker adds information in what is almost telegraphic speech: *Cost of a first class ticket Dallas to San Francisco departing August the 6th.*
- **Absence of additional connections:** in sentences like the following, the system cannot answer because *arrival times* are related to the flight relation and not to the fare ones in the relational database: *On fare code 7100325 list your arrival times from Dallas to San Francisco on August 14th.*
- **System incompleteness at the time of the test:**
 - in the sentence *Is there a flight from Denver through Dallas Fort Worth to Philadelphia?* the connection was established to handle a *from-to* relation, but not a *through* relation.
 - length of the sentences such as *Display lowest price fare available from Dallas to Oakland or Dallas to San Francisco and include the flight numbers on which these options are available.* This is a common problem in many NL systems.

Other sentences remain unanswered due either to some contradictory meanings in the lexical items of the queries or to the design of the database.

7 conclusion

It is important to note that the commercial system was partially adapted in a reasonable amount of time. Nevertheless, the overall system has not reached a fully satisfactory level of performance due to various factors:

- NLI was developed to handle "well-formed" written English so its performance is expectably poor for processing spoken and spontaneous queries.

- too small an amount of data was used to train the system. It would be profitable to use a larger and broader amount of training data, such as that available at SRI, CMU, and MIT.
- More time needs to be spent on refining the database connections and concepts.

The system in its current state is not yet fully operational. We are in the process of improving its performance. The focus of this paper is to point out the questions that a natural language interface (here the NLI system) faces in tackling spoken language. The interest for Bell laboratories is in the integration of speech and natural language. This is a difficult problem; in the context of the current architecture of the speech recognizer, the main question consists of processing the output of the recognizer to avoid too many multiple choices as input into the NLI component.

References

- [1] Bates, M., S Boisen and J. Makhoul, "Developing an Evaluation Methodology for Spoken Language Systems", *Proc. DARPA Speech and Natural Language Workshop*, 102-108, Hidden Valley, Pennsylvania, June 1990.
- [2] Bates, M., R. Robrow, S Boisen, R. Ingria, D. Stallard, "BBN ATIS System Progress Report - June 1990", *Proc. DARPA Speech and Natural Language Workshop*, 125-126, Hidden Valley, Pennsylvania, June 1990.
- [3] Bly B., P. J. Price, S. Park, S. Tepper, E. Jackson, V. Abrash, "Designing the Human Machine Interface in the ATIS Domain", *Proc. DARPA Speech and Natural Language Workshop*, 136-140, Hidden Valley, Pennsylvania, June 1990.
- [4] Ginsparg, J. M. *Natural Language Processing in an Automatic Programming Domain*, Ph. D. Dissertation, Stanford University, California, 1976.
- [5] Hirschman, L., D. A. Dalh, D. P. McKay, L.M. Norton and M. C. Linebarger, "Beyond Class A: A Proposal for Automatic Evaluation of Discourse", *Proc. DARPA Speech and Natural Language Workshop*, 109-113, Hidden Valley, Pennsylvania, June 1990.
- [6] Moore R., D. Appelt, J. Bear, M Darymple, and D. Moran, "SRI's Experience with the ATIS Evaluation", *Proc. DARPA Speech and Natural Language Workshop*, 147-150, Hidden Valley, Pennsylvania, June 1990.
- [7] *NLI Reference Manual*, Natural Language Incorporated, Berkeley, California, 1990.
- [8] *NLI Developer's Reference*, Natural Language Incorporated, Berkeley, California, 1990.
- [9] Norton L. M., D. A. Dalh, D. P. McKay, L. Hirschman, M. C. Linebarger, D. Magerman, and C. N. Ball, "Management and Evaluation of Interactive Dialog in the Air Travel Domain", *Proc. DARPA Speech and Natural Language Workshop*, 141-146, Hidden Valley, Pennsylvania, June 1990.
- [10] Pieraccini R., E. Levin, and C. H. Lee, "Stochastic representation of conceptual structure in the ATIS task", *Proc. DARPA Speech and Natural Language Workshop*, Asilomar, California, June 1991.
- [11] Ward W., "The CMU Air Travel Information Service: Understanding Spontaneous Speech", *Proc. DARPA Speech and Natural Language Workshop*, 127-129, Hidden Valley, Pennsylvania, June 1990.
- [12] Zue, V., J. Glass, D. Goodine, H. Leung, M. Phillips, J. Polifroni and S. Seneff, "Preliminary ATIS Development at MIT", *Proc. DARPA Speech and Natural Language Workshop*, 130-135, Hidden Valley, Pennsylvania, June 1990.