**WHEELS: A CONVERSATIONAL SYSTEM IN THE AUTOMOBILE CLASSIFIEDS DOMAIN**

Helen Meng, Senis Busayapongcha†, James Glass, Dave Goddeau, Lee Hetherington, Ed Hurley, Christine Pao, Joe Polifroni, Stephanie Seneff, and Victor Zue

Spoken Language Systems Group
MIT Laboratory for Computer Science
Cambridge, MA 02139 USA
BellSouth Information Systems †
Tucker, GA 30084 USA
http://www.sls.lcs.mit.edu

**ABSTRACT**

WHEELS is a conversational system which provides access to a database of electronic automobile classified advertisements. It leverages off the existing spoken language technologies from our GALAXY system, and enables users to search through a database of 5,000 automobile classifieds. The current end-to-end system can respond to spoken or typed inputs, and produces a short list of entries meeting the constraints specified by the user. The system operates in mixed-initiative mode, asking for specific information but not requiring compliance. The output information is conveyed to the user with visual tables and synthesized speech. This system incorporates a new type of category bigram, created with the innovative use of the natural language component. Future plans to extend the system include operating in a displayless mode, and porting the system to Spanish.

1. **INTRODUCTION**

Over the past year we have been developing a conversational system named WHEELS, interfacing to a database of electronic automobile classified ads. The ultimate goal of this work, done in conjunction with BellSouth Intelliventures, is to develop a displayless, bilingual (English and Spanish) conversational system that can be deployed over the telephone network. We plan to utilize the context of this application to experiment with new speech technologies. The intermediate goals of this project, from a system development perspective, are (1) to port our GALAXY framework to the WHEELS domain, using telephone speech input, (2) to develop a displayless version of the conversational system, including an interface with the narrow-band recognizer for telephone-based displayless deployment, and (3) to extend the system to Spanish. We have completed the first stage, and are beginning development of the displayless system. At present, our end-to-end system is capable of responding to speech or typed inputs, and produces a short list of entries meeting user-specified constraints (e.g., in make, model, price, year, mileage, etc.). The output information is conveyed to the user via a visual table display and synthesized speech.

The following section describes the various components of our system. Section 3 describes our experiments in generating a category bigram from a context free grammar. In Section 4 we discuss the data collection efforts done in conjunction with BellSouth Intelliventures. A results section reports on recognition performance and parse coverage. We conclude with our future plans.

2. **SYSTEM COMPONENTS**

Figure 1 shows a block diagram of WHEELS. This system preserves the GALAXY client-server architecture [3]. The GALAXY client interfaces between the user and the system. It also contains TINA [6] and GENESIS [1] libraries which are used for natural language understanding and language generation respectively. GENESIS is utilized to paraphrase the input sentence, so as to reflect the interpretation of the input by the system, serving to alert the user to any recognition errors. The WHEELS server is responsible for discourse and dialogue modeling as well as database retrieval. It makes use of GENESIS for its response generation. A detailed description of the dialogue component can be found in [4]. The system utilizes the FastFind database, provided to us by BellSouth Intelliventures. It contains about 5,000 entries of automobile classified advertisements that appeared in the Atlanta Journal and Constitution.

![](image1.png)

Figure 1: System Architecture of WHEELS.

2.1. **Speech Recognition**

The speech recognition component utilizes a stochastic segment-based recognizer developed in our group called SUMMIT [5]. Context independent phone models were initially trained on the ATIS sentence subset of the MACROPHONE corpus obtained from the Linguistic Data Consortium. They were later retrained on a large corpus.

---

† This research was supported by a research contract from BellSouth Intelliventures.
of domain specific utterances collected in a wizard mode (see Section 4). It makes use of a category bigram that was derived using a new approach based on our natural language system, as described in the following section. The vocabulary contains over 1,100 words. A more complete description of the recognizer is available in [2].

2.2. Natural Language Understanding

Each of the \( N \)-best recognition hypotheses \((N=10)\) produced by SUMMIT is processed by the natural language parser, TINA. TINA parses with a a context-free grammar, augmented with syntactic and semantic features, as well as a trace mechanism to handle gaps [6]. Currently the WHEELS grammar consists of approximately 590 non-terminal categories (which are mixed syntactic and semantic categories), and 920 terminals, supporting about 66 auto makes, 520 models and 50 submodels. The highest scoring recognition hypothesis which produces a full parse is converted by TINA into a semantic frame. Figures 2 and 3 show the parse tree and the corresponding semantic frame for the sentence “SHOW ME BLACK CHEVROLET CORVETTE CONVERTIBLES.” If the parser fails to find a full parse in the top \( N \) sentences, the semantic frame is generated from the highest-scoring robust parse.

![Figure 2: Parse tree for the sentence “SHOW ME BLACK CHEVROLET CORVETTE CONVERTIBLES.”](image)

![Figure 3: Semantic frame for the sentence “SHOW ME BLACK CHEVROLET CORVETTE CONVERTIBLES.”](image)

| sentence \( \rightarrow \) [word*1] np_auto [word*2] np_auto [word*3] |
| np_auto \( \rightarrow \) make | model     | make_model | auto |
| make_model \( \rightarrow \) honda accord     | ford ranger |
| auto \( \rightarrow \) sports utility (van | vehicle...) |

Figure 4: Representative entries from the category bigram rules file. "|" indicates alternatives; [ ] indicates optional.

2.3. Dialogue Modeling

The WHEELS domain server has two main responsibilities: to manage the dialogue interaction, and to retrieve the appropriate items from the database. The user can specify cars by make, model, year, price, mileage, and other features such as color, place of origin, car-type (e.g., sport utility vehicle), etc. The semantic frame generated from the user’s spoken input is received by the WHEELS domain server, and cast into the current dialogue context. This is accomplished via a “form-filling” mechanism, where the information from the semantic frame is entered into the appropriate “slot” of an electronic form (or “E-form”), which is a structure for maintaining coherence in each dialogue [4]. All the information present in the E-form is used to generate an SQL query to access the database.

In addition, the system prompts the user for further constraints to narrow down the search (e.g., “A price range would be helpful.”), based on information absent from the E-form. It can subsequently accommodate both a cooperative response (which answers the prompt directly) and a non-cooperative response (which ignores the prompt). Such flexibility is inherent in the design of the discourse model. Further details regarding dialogue modeling in WHEELS can be found in a companion paper [4].

3. DERIVED CATEGORY BIGRAM

Whenever a new domain is first developed, there is always the problem of insufficient training data for the recognizer’s language model. This problem can potentially be overcome by deriving a category bigram from the grammar rules that are being developed in parallel for the natural language component. Such a bigram has the advantage that it can be well matched to the constraints of the natural language component.

The basic principle, as applied here, is to use a simple context-free grammar to make explicit any obvious patterns in the domain, and cover the rest of the words through a flexible all-word rule. To this end, we have developed the capability of expressing a flexible grammar that intermixes a word* rule with a set of recursive context-free rules that cover the prominent word patterns in the domain. The recursive rules specify, for example, various ways of naming cars, as in “Chevrolet Corvette convertible.”

Representative rules for this grammar are given in Figure 4. The rules show three unique word* labels, tagged positionally relative to an np_auto category. Although all three instances have the same all-word model, their probability distributions will differ significantly, due to their differing positions in the sentence. These differences yield significant improvements in the perplexity of the result-

---

2 Augmented with obvious categories such as color and digit.
3 For example, we expect that the category word*2 would consist mostly of conjunctions, e.g., “I'D LIKE HONDA CIVICS AND PRELUDES.”
Two-category bigram

Bigram

Standard

understanding model, over what would be realized with a single shared \textit{word*} model.

In order to produce a preliminary category bigram from such a grammar, whatever small amount of training utterances available initially are parsed using this covering grammar. Rule counts are tabulated from parsed sentences and converted to conditional sibling-sibling probabilities, which can then be multiplied out to the preterminal level, yielding a standard category bigram. For high perplexity points where sparse data are likely to be problematic, the frequency counts can be augmented with estimates obtained from other sources, such as the frequency of occurrence in the database. Such database-driven frequency counts are spliced into the rules for make and model, giving strong preference to the Fords and Hondas, for example.

This strategy has been found to be useful in the \textsc{wheels} domain. It continued to be beneficial even after we had obtained a large amount of training data (see Section 6). We believe that the main advantage of the derived bigram technique in \textsc{wheels} is its ability to generalize across all references to make-model pairs in cars, with respect to the external context in the sentence, while at the same time preserving the strong ties between particular makes and their associated models. This leads directly to better recognition of make-model pairs. A standard category bigram, with the obvious \textsc{make} and \textsc{model} categories, would not be able to enforce the constraint that a “Ford Taurus” is far preferred over a “Ford Civic.” An undesirable alternative would be to explode the vocabulary with tied words enumerating all legitimate make-model pairs, but this would lead to far greater inefficiencies in the recognizer.

### 4. DATA COLLECTION

During the fall of 1995 BellSouth has used the \textsc{wheels} system to collect domain-specific data from two cities in Florida – Jacksonville and Melbourne. The data collection system was set up in two shopping malls, and therefore our subjects involve real users interacting with our system in a realistic environment. We have obtained over 14,000 utterances from about 600 speakers (male-to-female ratio was approximately 2:1). A wizard typed user queries directly into the \textsc{wheels} system, and speech was recorded from the telephone handset. Subjects were requested to continue the dialogue until the session was complete. In many cases, the subject chose to provide us with multiple dialogue sessions. The domain-specific data have been used for training our speech recognizer, broadening the coverage of the \textsc{tina} grammar, refining the \textsc{wheels} domain server, and offering guidance for the design of the dialogue component.

### 5. EXAMPLE DIALOGUES

Figures 5 and 6 illustrate the capabilities of \textsc{wheels}. Figure 5 shows an interaction between the system and a cooperative speaker, who always responds directly to the system’s prompts. Figure 6 is an exchange between the system and a non-cooperative speaker, who often ignores the system’s prompts. A dialogue session is considered complete when the user successfully narrows the search to five or fewer cars.

BellSouth has been involved in the development of an Interactive Voice Response (IVR) application thus far restricted to requests for make and model, intended to replace a touch-tone input. In order to

### Table 1: Perplexity and recognition results for the regular and category bigram

<table>
<thead>
<tr>
<th>Data set</th>
<th># utts</th>
<th>Bigram</th>
<th>Perplexity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>1200</td>
<td>standard</td>
<td>9.1</td>
<td>67.4%</td>
</tr>
<tr>
<td>Test</td>
<td>1200</td>
<td>category</td>
<td>9.9</td>
<td>69.3%</td>
</tr>
<tr>
<td>IVR (make-model)</td>
<td>383</td>
<td>standard</td>
<td>6.1</td>
<td>74.8%</td>
</tr>
<tr>
<td>IVR (make-model)</td>
<td>383</td>
<td>category</td>
<td>5.9</td>
<td>76.3%</td>
</tr>
</tbody>
</table>
assess the feasibility of such a system, we decided to measure the recognition performance using another disjoint data set of over 380 sentences which are limited to make and/or model requests. For this limited subdomain it is possible to measure “task completion” success rate. Our criteria for success are tabulated in Table 2.

As shown in Table 1, perplexity measured using the IVR make-model test set was 6.1 for the standard bigram and 5.9 for the derived bigram. A comparison between the perplexity measurements of the two test sets may suggest that the derived category bigram can better capture the constraints between makes and models, but this effect may be outweighed by constraints lost in the all-word modeling between categories. Recognition accuracies show consistent improvement between the two test sets as we move from the standard bigram to the derived bigram. The 74.8% word accuracy using the standard bigram translates to a task completion accuracy of 79.1%, while the 76.3% word accuracy using the derived bigram translates to a task completion accuracy of 79.4%. In the actual application, it is intended that the IVR system will verify the top few hypotheses through user feedback. If there is a positive verification, the task is considered complete. Otherwise, the user is asked to spell out the make and/or model of the car, and only the top-scoring spelling hypothesis is considered. Preliminary experiments indicate that augmentation with the spelling confirmation procedure would significantly boost the task completion accuracy to over 90%.

7. FUTURE WORK

The near-term goals with regards to the WHEELS project involve moving the development towards a displayless mode of operation, and porting the system to Spanish. In addition to general telephony issues that need to be addressed for a displayless application, it will present interesting challenges in both language generation and discourse management. The content of the classified ads will have to be delivered succinctly to the user; there will also be additional subdialogs required for clarification, requesting information to be read or repeated, keeping track of what fraction of a search list has been read, referring to and inquiring about specific items in the list, as well as discourse-dependent help instructions. Porting WHEELS to the Spanish language will uncover its similarities and differences compared to English for parsing and generation purposes, and promote portability across languages.

8. ACKNOWLEDGEMENT

We would like to acknowledge the assistance and expertise of our collaborators at BellSouth Intelliventures, especially Dewey Anderson and Dave Anderson, for their help in coordinating the project and collecting the large body of field data in Florida.

9. REFERENCES