Experimental Results

We evaluated the preliminary system on three Pilot Tests. The main purpose of these tests is to reduce the Out Of Vocabularies (OOVs) for both Natural Language Parser and ASR. Both ASR and NL Parser get the 0% OOVs after the three tests. The test performance is summarized below in Table 1.

The reason for the Parser's error rate does not go to 0% is that the parsing results are not necessarily understood by Dialog manager and sometimes lack enough information for Dialog Manager.

Table 1: Results of Pilot Test 1, 2, 3

Database	1	1.1	2	2.1	3
LM Processing					
(a) OOV Rate	67/92	0/92	24/80	2/80	17/104
(b) WSJ PPL	115	107	107	108	108
(c) Test PPL	253	399	406	190	223
(d) Lex wsj/cavs	4988/6083	4988/7408	4988/7408	4988/7111	4988/7111
(e) WSJ AvgScores	-2.8 ~ -8.2	-2.8 ~ -8.2	-2.8 ~ -8.2	-2.8 ~ -8.2	-2.8 ~ -8.2
(f) Cavs AvgScores	-	-4.4~-7.8	-4~-7.5	-3.2~-6.8	-3.4~-7.8
NL Processing					
(g) OOV Rate	23/92	0/92	29/80	0/80	5/104
(h) Parser	75/92	3/92	49/80	4/80	48/104

Then, we tested the overall system performance over the three pilot databases.

Table 2: Results of Overall System Test

Query Category	ASR Processing		NL Processing Parser Error Rate	Dialog Manager		Database Query Error Rate
	Word Error Rate	Sentence Error Rate		Understanding Error Rate	Response Error Rate	
Test 1 (Feb. 3)	N/A	N/A	121/274	137/274	274/274	274/274
Test 2 (Feb. 24)	N/A	N/A	19/274	20/274	140/274	135/274
Test 3 (Mar. 04)	42.2%	242/274	213/274	213/274	229/274	229/274
Test 3.1 (Mar. 15)	40.3%	238/274	N/A	N/A	N/A	N/A
Test 3.2 (Mar. 15)	43.0%	227/274	N/A	N/A	N/A	N/A

Notes for Test 1:

- The error rate of Parser is high because it can not differentiate the query types. It doesn't recognize whether the question is asking for direction, address and building. So we create more different semantic frames to understand this.
- Dialog Manager needs to be modified to deal with these different new semantic frames.
- The communication interface between DM and DB needs to be defined.
- Test 1 is conducted using the old Drive Frame structure for Parser, that means only one frame (drive) in table 4 exists at the time of testing, but the Drive Frame was planned to take all the queries at that time.

Notes for Test 2:

- After we created more semantic frames to handle different types of queries, the error rate of Parser went down to 19/274.
- We have not found a proper database source for campus-information query. Almost half of query errors of Database were resulted from this.
- Dialog Manager and Parser need more intelligently deal with words like "nearest", "closest" - translating them to query "by distance".
- Database died a lot during test, which was very annoying.
- Test 2 is conducted after Drive, Query_Address, Query_Direction and Query_Building frames are implemented except the query related to campus names.
- NL Processing Parser Error Rate means the number of utterances which can not be

understood by the Parser.
- DM Understanding Error Rate means the number of utterances which can not understood by the DM.
- DM Response Error Rate means the number of utterances which can not presented correctly by the DM.
- DB Query Error Rate means the number of answers to utterances which can not found by the DM.

Notes for Test 3:

- As we can see, most understanding errors are resulted from the recognition errors.

Here is a list of existing problems for the system.

Table 3: Outstanding Problems for Ongoing Components

Servers	Problems
ASR	OOVs
	LM
	Pronunciations
	Efficiency
	WER
	Noise Robustness
	Training data
NLP	OOVs
	Grammar rules
	Sentence structure
Dialog Manager	Semantic frames
	Dialog strategies
Database Server	Data Resources

Here is a list of existing sentence type vs. query type for the system.

Table 4: Sentence classification vs. Semantic Frame

Sentence Type	Query Type	Sub Type	Status	Example(s)
General sentence without any goal (can not be classified into any following cluster)	[Drive]	General	finished	I'm at Wendy's.
query address campus address	[Query_Address]	Place name	finished	Where is Walmart?
		Name near Name	finished	Where is McDonald's near Kroger?
		Name near Address	finished	Where is McDonald's near Highway 25?
		Place Specification + Place Type	finished	List all Chinese restaurants in starkville?
		Campus Name	pending	Where is MCCAIN? Where is the dean's office
query direction query distance campus direction	[Query_Direction]	From Name To name	finished	How can i drive from walmart to foodmax?
		From (to) name near name	implementing	How can i drive to china garden near kroger?
		From address	finished	How can i

		to address		drive from lee blvd to main street?
		From (to) name near address	implementing	How can i drive to mcdonald's near highway 25?
		name	finished	How far is china garden?
		From (to) name near name	implementing	How far is mcdonald's near walmart?
		From (to) name near address	implementing	How far is mcdonald's near highway 25?
		Campus name	pending	Can you help me to get from simrall to butler?
query building	[Query_Building]	General	finished	Where building is ece in?
query turn	[Query_Turn]	General	pending	I'm on Nash street what's my next turn
query bus	[Query_Bus]	General	pending	What shuttle bus routes stop at the coliseum
query intersection	[Query_Intersection]	General	pending	Does Lynn Lane intersect Academy

				Road?
query quality	**[Query_Quality]**	**General**	**pending**	**How can I find a good restaurant nearby? Find the most scenic path from LJ's to Scott Field**
query special (Some difficult queries to this cluster)	**[Query_Special]**	**General**	**pending**	**Can I bypass highway 12 to get to Bryan field? Where can I buy some souvenirs in Starkville?**

Here, we gradually improve the system performance over the five action items.

Table 5: Results of Action Items

Action Item	Updated Date	Progress
(I) Improve the performance on current Drive task	Feb. 8	Finished first part of [query_direction] such as "where is wendy's?"
	Feb. 25	See "**status**" column in Table 4
(II) Cluster all queries by the type of database they need	Feb. 8	Several types of database are necessary: 1. Web sites such as mapquest.com 2. Local database A: campus building names to academic names: **Finished** 3. Local database B: campus distance queries: **Investigating** 4. Local database C: campus's location query: **Investigating** 5. Local database D: quality query (such as good, best): **Investigating**
	Feb. 25	Worked on implementation and improvement of databse 1, 2, which made the database query error rate go down to 135/276 from 276/276 in Table 2.
(III) Determine the semantic frames we need	Feb. 8	[**query_address**]: where is wendy's? [**query_direction**]: how can i drive from holiday inn to wendy's? [**query_location**]: which building is the ECE department in? [**query_distance**]: how far? [**query_bus**]: bus route or whatever. [**query_others**]: some hard questions but may not be queried.
	Feb. 25	See Table 4
(IV) Principle way to improve the parser and to access the performance	Feb. 8	(III) gives a good way to improve Parser performance and make the better understanding for Dialog Manager
	Feb. 25	By implementing the multiple semantic frames described in (III) to handle different types of queries, the error rate of Parser went down from 137/276 to 19/276 in Table 2
(V) Improve grammar rules to handle pilot tests	Feb. 8	Added grammar rules for each semantic frames and finished [query_address] and [query_direction].
	Feb. 25	Added grammar rules for [query_address], [query_direction], [query_distance] and [query_building].