

Conversational Re-prompting in Natural Language Dialog

Harry E. Blanchard & Osamuyimen T. Stewart
AT&T Shannon Labs
Middletown, NJ, USA

hblanchard@research.att.com

ustewart@att.com



HFES Conference – Sep 23, 2004

Overview

- Natural language dialog
 - AT&T VoiceTone™ How May I Help You?SM Technology
- First turn dialog
- Second turn dialog
- Re-prompting
 - Caller asks for representative
 - Caller is vague
- Comparison of re-prompting dialog strategies
- Results from implementation in on-line production system
- Conclusions and future research

Speech Technology Systems for Telephony

IVR = Interactive Voice Response

- Technology continuum: Machine's ability to respond
 - Touch-tone
 - Command word and phrase
 - Slot Filling
 - Natural Language

- Example – directed dialog

For your account balance, say 'balance', if you want to know our mailing address, say 'address', to pay your bill by credit card, say 'pay bill'

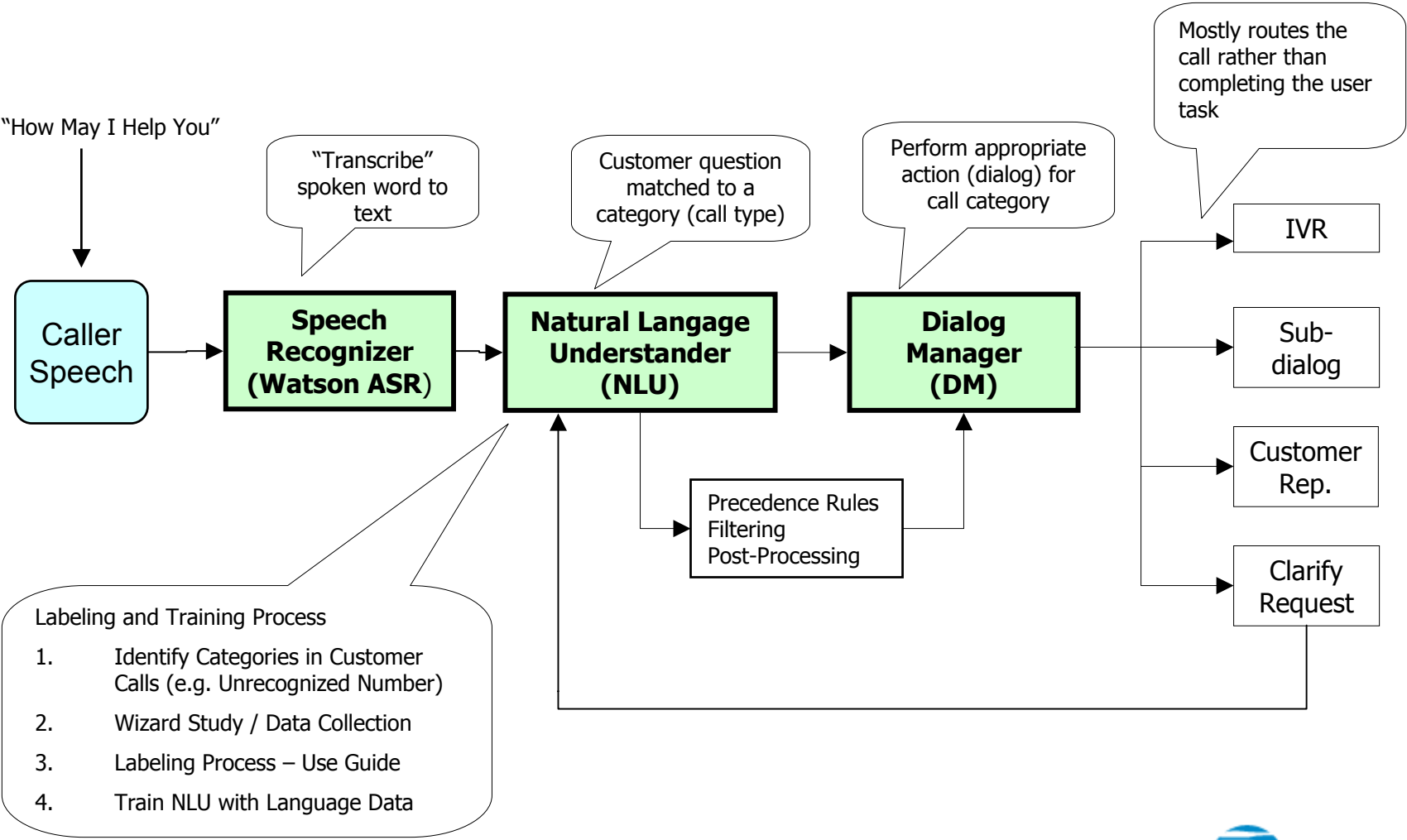
Natural Language Understanding

- Different from 'directed dialog' ASR Systems:
 - No predetermined words or phrases which the caller must say to choose an option
 - Users ask a question freely in their own words
 - No vocabulary or syntax constraints on natural speech
- Example – natural language dialog
 - *Hello, I am an automated assistant. You can speak naturally to me. How may I help you?*
 - *Um, I'd like to use my credit card to pay this month's bill*
 - System identifies category as PAY BILL
 - Rejects ACCOUNT BALANCE, MAILING ADDRESS, etc.
 - *Okay, I can help you with that.*
 - *Please wait while I transfer you to the billing system.*

Natural Language Understanding

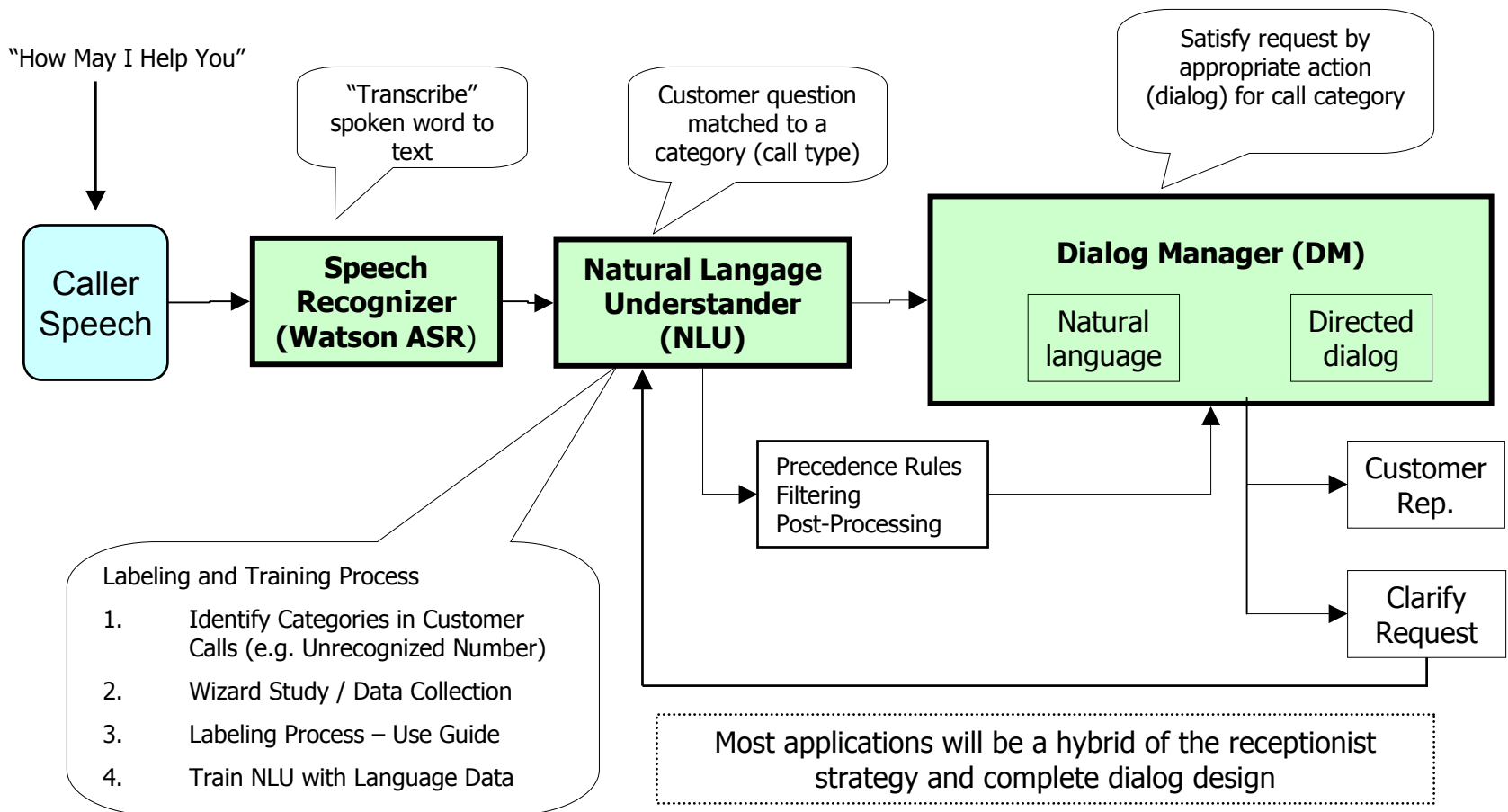
- Data Driven
 - Machine categorization
 - Categories of requests are defined (based upon data)
 - PAY BILL, ACCOUNT BALANCE, MAILING ADDRESS, etc.
 - Examples from human-human or human-machine conversations are harvested and used to 'train' NLU system
- Run-time system
 - Prompt does not specify exact phrase
 - Open-ended, conversational
 - Suggest general categories,
 - Suggest general phrases
 - User responds in unconstrained manner
 - System classifies all responses using NL technology

HMIHY (Receptionist Design)

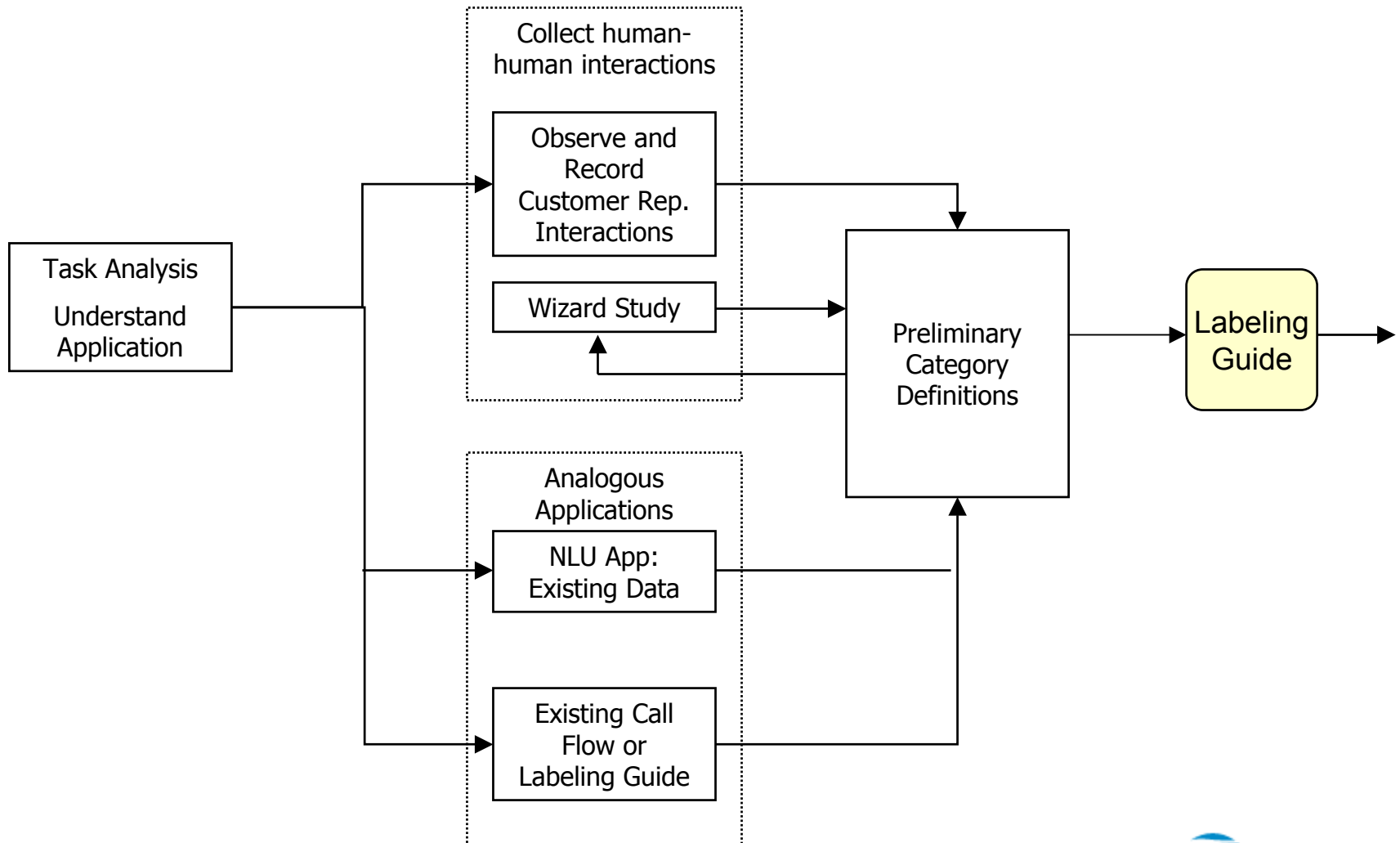


HMIHY – Complete Dialog Design

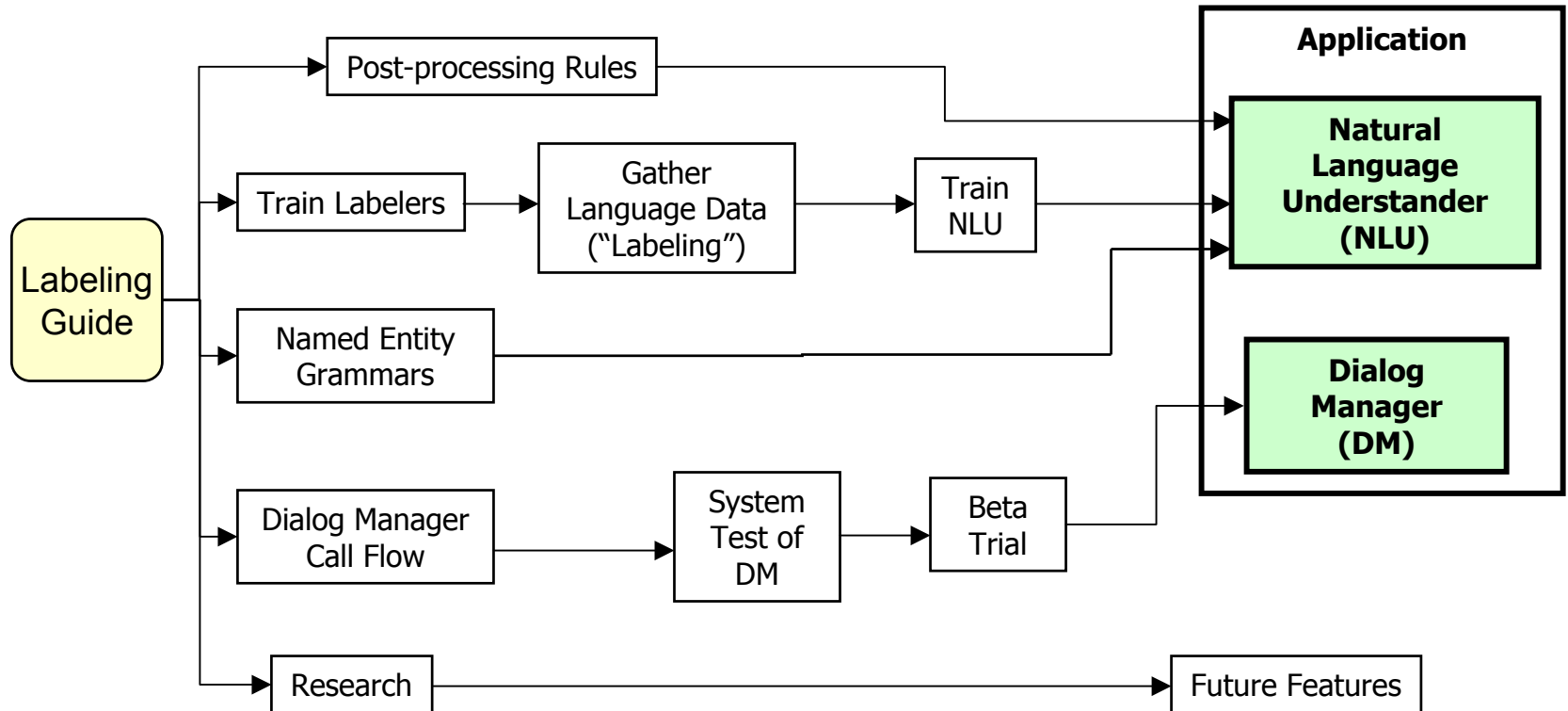
The caller's task is dealt with entirely by the dialog manager



Language Analysis and Category Identification

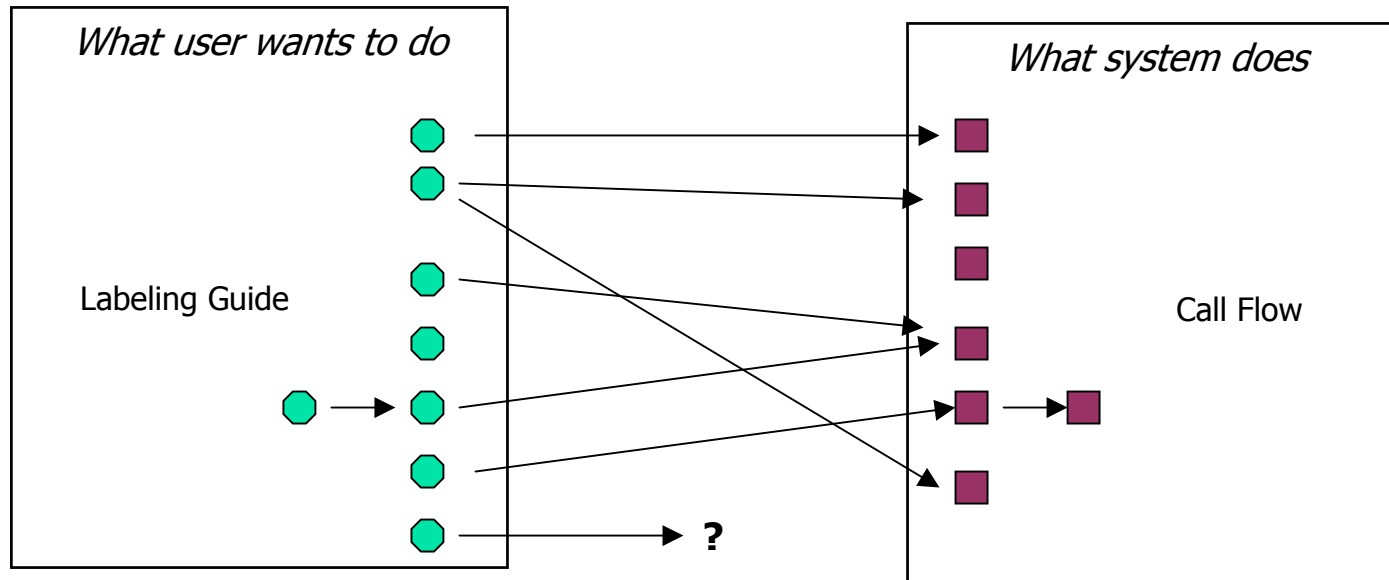


Labeling Guide Drives NLU Training and Dialog Manager Design



Mapping of User Intent to System Capabilities

Relationship Between Labeling Guide and Call Flow Design



Example: Consumer Telecom Customer Care – Unrecognized Number and Credit

- User will ask about unrecognized number – intention is to get credit for unrecognized number
- User will ask for credit – wants to check on unrecognized number before negotiating for credit

Advantages to the User Experience with Natural Language Understanding

- The *customer* controls the conversation
 - This is a major reason why HMIHY increases customer satisfaction
- The customer uses their *own language*
 - No need to recognize and understand company terminology
 - The customer focuses more on what they want rather than spending attention and memory on options and keywords for all possible tasks
 - Many customers will spend less time search and exploring thru an IVR
- Receptionist and complete dialog models are possible
 - Customers are able to flexibly change dialog styles with the receptionist
- More customer service tasks could be automated in HMIHY than touchtone or ASR-based IVR ... *without loss of usability*
 - IVR must get more and/or bigger, complex menus or more keywords
 - HMIHY will just recognize more types of customer requests

User Interface Design in Natural Language

- Lends itself to conversational and open-ending prompting
 - Performance of technology is linked to natural requests
- Understanding and modeling of human conversation is essential
 - Category identification and labeling from language data
 - Understanding of conversation essential to dialog design
- Design of prompts can be crucial to system performance
 - Seemingly simple changes can alter caller language significantly
- Proper handling of errors and lack of progress in conversation is crucial to system success
 - The constraint is that users must express a specific, identifiable request

First Turn Prompt

- Boyce, 2000 – empirical evaluation of NLU greeting prompts
- Long versus short prompts:
 - *How may I help you?*
 - *I am an AT&T automated assistant. You can speak naturally to me. How may I help you?*
- Variations in the greeting prompt significantly effected performance of the NLU, because it significantly altered language behavior
 - Variations primarily effect the responses in turn 1 (which makes sense).
 - The longer prompt results in the highest percentage of callers giving a specific call type in the first turn and has the lowest no-information responses.

Other Strategies for First Turn

- “Priming” Sheeder & Balogh (2003)
 - Placed examples of responding in greeting prompt before or after HMIHY question
 - Specific keywords
 - Example phrases of natural questions
 - Example - priming
Welcome to Clarion Wireless Customer Service. You can ask me things like “how many minutes have I used?,” and “I’d like to set up automatic payments.” So, how can I help you with your account?
- Results
 - Good results in experimental task (Sheeder & Balogh, 2003)
 - Poor results in the field (AT&T VoiceTone deployments)
 - Users repeated examples
 - Users unclear that there are more options than in examples
- *Alternative: Modeling of human conversation -- suggest concepts rather than keywords or example phrases*
 - *End Focus Principle (Quirk et al., 1985)*

Second Turn Prompt

- Dialog may fail on first turn in several ways:
 - The caller says something that is not relevant to the task
 - "Are you a machine?" -- "Nice day" etc.
 - Task-relevant but too vague to classify
 - "I'm having a problem" -- "I have a billing question" -- "I need something" etc.
 - The caller wants to opt out of automation immediately
 - Asks to talk to a person -- "I need to talk to a real person"
 - Silence, noise, or other situation which cannot be understood
- Re-prompt must encourage more specific task-relevant request on second turn
- Too many unclassifiable responses → leave automation
 - Typically: third turn

Alternative to Priming: Second Turn Re-prompt

- Open ended greeting
 - Natural and efficient interaction for users who respond and system understands
- Re-prompt only with explanations, examples, etc.
 - Use more aggressive, directed prompting only when there's a problem
- Model re-prompt on human conversation
 - Use concepts rather than keywords or examples

Field Study

- Working system with live customer traffic
 - Pharmaceutical industry end-consumer IVR
 - Customer care – prescriptions-by-mail
- Boyce (2000) style greeting prompt
 - Re-prompted if failed to get specific request on first turn
- Compared
 - General re-prompt (“Okay, what’s your question?”)
 - Re-prompts with more directed prompting and suggestions
- Examined two types of failures
 - Caller asks for a person
 - Caller is too vague

Caller Asks For a Person

- First turn: *Welcome to <service name>. I am an automated assistant. You can speak naturally to me. How may I help you?*
 - *I want to talk to a customer service representative*

- Second turn:
 - Before:
You can go ahead and speak naturally to me. I can understand your speech and help you get the service you need, just briefly tell me how I may help you.

 - After:
Okay. In order to direct your call please tell me if you need to refill an existing medication, find out the status of your prescription order, or anything else you want to speak to the customer service representative about.

Vague Requests

- Before
 - First turn: *Welcome to <service name>. I am an automated assistant. You can speak naturally to me. How may I help you?*
 - *I'm having a problem.*
 - Second turn: *Okay what's your question?*
 - *SPECIFIC RESPONSE* → "Okay I'll connect you"
 - *VAGUE RESPONSE* → "Okay, I'm going to connect you to a customer service representative ..."
- After
 - Use ability of NLU to extract partial information
 - Degrees of ambiguity exist:
 - Vague – only know it's task relevant "I have a problem"
 - Request may be about an order: order status, place an order, ...?
 - "I'm calling about an order."
 - Request may be about billing: bill dispute, pay bill, missing bill, ...?
 - "I have a problem with my bill."
 - Etc.



Vague Requests

- After
- Vague request about an order
 - First turn: *Welcome to <service name>. I am an automated assistant. You can speak naturally to me. How may I help you?*
 - *I'm calling about an order*
 - Second turn: *I'm sorry, do you need to refill an existing medication or get status on an order you've already sent in? Please tell me how I may help you.?*
 - *SPECIFIC RESPONSE → "Okay I'll connect you"*
 - *VAGUE RESPONSE → "Okay, I'm going to connect you to a customer service representative ..."*

Vague Requests

- All other vague requests
 - First turn: *Welcome to <service name>. I am an automated assistant. You can speak naturally to me. How may I help you?*
 - *I'm having some problems*
 - Second turn: *Okay, what's your question?*
 - *Caller → specific or vague response*
 - Third turn:
 - *SPECIFIC RESPONSE → "Okay I'll connect you"*
 - *VAGUE RESPONSE → "I'm sorry, do you need to refill an existing medication or get status on an order you've already sent in? Please tell me how I may help you.?"*
 - Fourth Turn
 - *Caller → specific or vague response*
 - *SPECIFIC RESPONSE → "Okay I'll connect you."*
 - *VAGUE RESPONSE → "Okay, I'm going to connect you to a customer service representative ..."*

Results

- Examined final resolutions
 - *Last state prior to fulfilling request or transferring to a person*
- Measured % of final resolutions that are
 - Desired = specific request
 - Undesired = vague, or ask for a person
- Comparison 1: Requests for a person
 - Final Resolution = ask for a person
 - (Asks for a person on first and second turn)
 - Before = 2.5 %
 - After = 0.6 %

Total calls

- Before = 32,221
- After = 46,423

Results

- Comparison 2: Vague Requests
- Final Resolution = vague
 - (Vague on first and second turn)
 - Before = 2.7 %
 - After = 2.6 %
- Vague request about an order only
 - Final resolution = vague
 - Before = 0.44 %
 - After = 0.33 %

Conclusions

- Avoid more directed prompting in first prompt
- Re-prompt with more specific information about types of requests
 - Increased success rate
 - Did not sacrifice conversational nature and naturalness of dialog
- Can use partial information from user's response to have tailored, specific re-prompt