Error Handling in Spoken Dialogue Systems

Svetlana Stoyanchev
Seminar on Spoken Dialogue Systems
03/09/2015

Outline

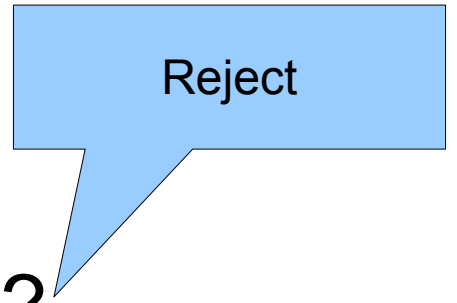
- How do human speakers handle errors in a dialogue
- How do dialogue systems handle errors
  - Overview of dialogue system strategies
- User reactions to system error recovery attempts
- New proposed approach to handling errors

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?



Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?

NLU clarification

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?
- (c) Did BO leave?

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?
- (c) Did BO leave?

Confirmation

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?
- (c) Did BO leave?
- (d) Did WHO leave?

Clarification

Human Clarification Strategies

- A: Did Bo leave?
- B:
- (a) I'm sorry, what did you say?
- (b) Who do you mean by `Bo'?
- (c) Did BO leave?
- (d) Did WHO leave?
- (e) Bo?
- (f) Who?

Clarification in human dialogue: Wh question

- A: No he's, he's being moved to troop fifteen
- B: To where?
- A: Troop fifteen
- B: Oh

Clarification in human dialogue: gap

- A : Can I have some toast please?
- B: Some?
- A: Toast

- Humans tend to ask targeted clarification questions

Outline

- How do human speakers handle errors in a dialogue
- How do dialogue systems handle errors
  - Overview of dialogue system strategies
- User reactions to system error recovery attempts
- New proposed approach to handling errors

Human-Computer Dialogue

SYS:	Welcome to the CMU Let's Go bus information. What can I do for you?
USR: ASR:	I need to go from Oakland *I need to go .from. can't*
SYS:	Where do you wanna leave from?
USR: ASR:	to Waterfront *told. .me. Waterfront*
SYS:	Leaving from Waterfront. Did I get that right?
USR: *ASR:*	Oakland *Oakland*
	…

Interaction is impaired by ASR errors

Component responsible for handling errors

Error Recovery Strategies

- Error detection is probabilistic
- Use confidence threshold of the ASR/NLU system to distinguish:
  - Potential misunderstanding:
    - System has a hypothesis of a user's concept
  - Potential non-understanding:
    - No hypothesis

Error Recovery Strategies: Misunderstanding

Error Handling Strategy (misunderstanding)	Example
Explicit confirmation (CLARIFY*)	Leaving from Waterfront, did I get that right?
Implicit confirmation (DISPLAY *)	Leaving from Waterfront … where are you going to?

*Skantze 2007 terminology

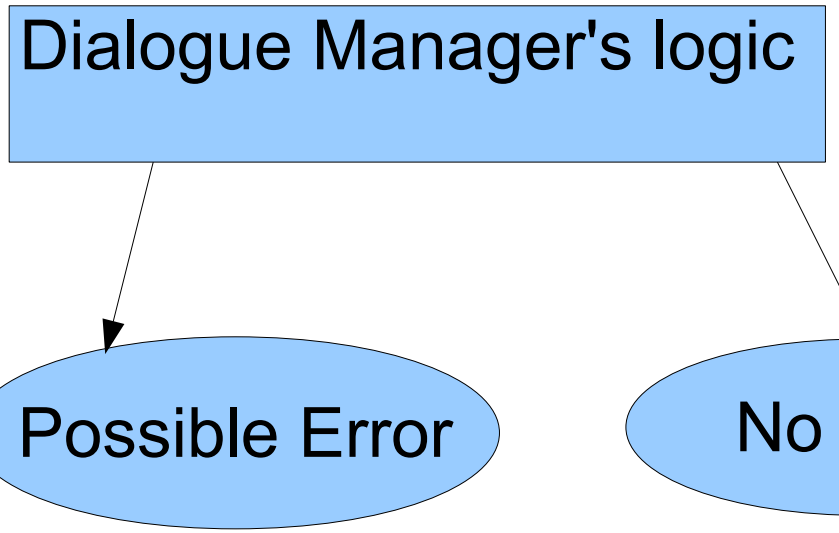
18

Error Recovery Strategies: non-understanding

Error Handling Strategy (non-understanding/rejection)	Example
Notify that a non-understanding occurred	Sorry, I didn't catch that .
Ask user to repeat	Can you please repeat that?
Ask user to rephrase	Can you please rephrase that?
Repeat prompt	Where are you leaving from?
Help Message	You can say "I am leaving from Downtown"

19

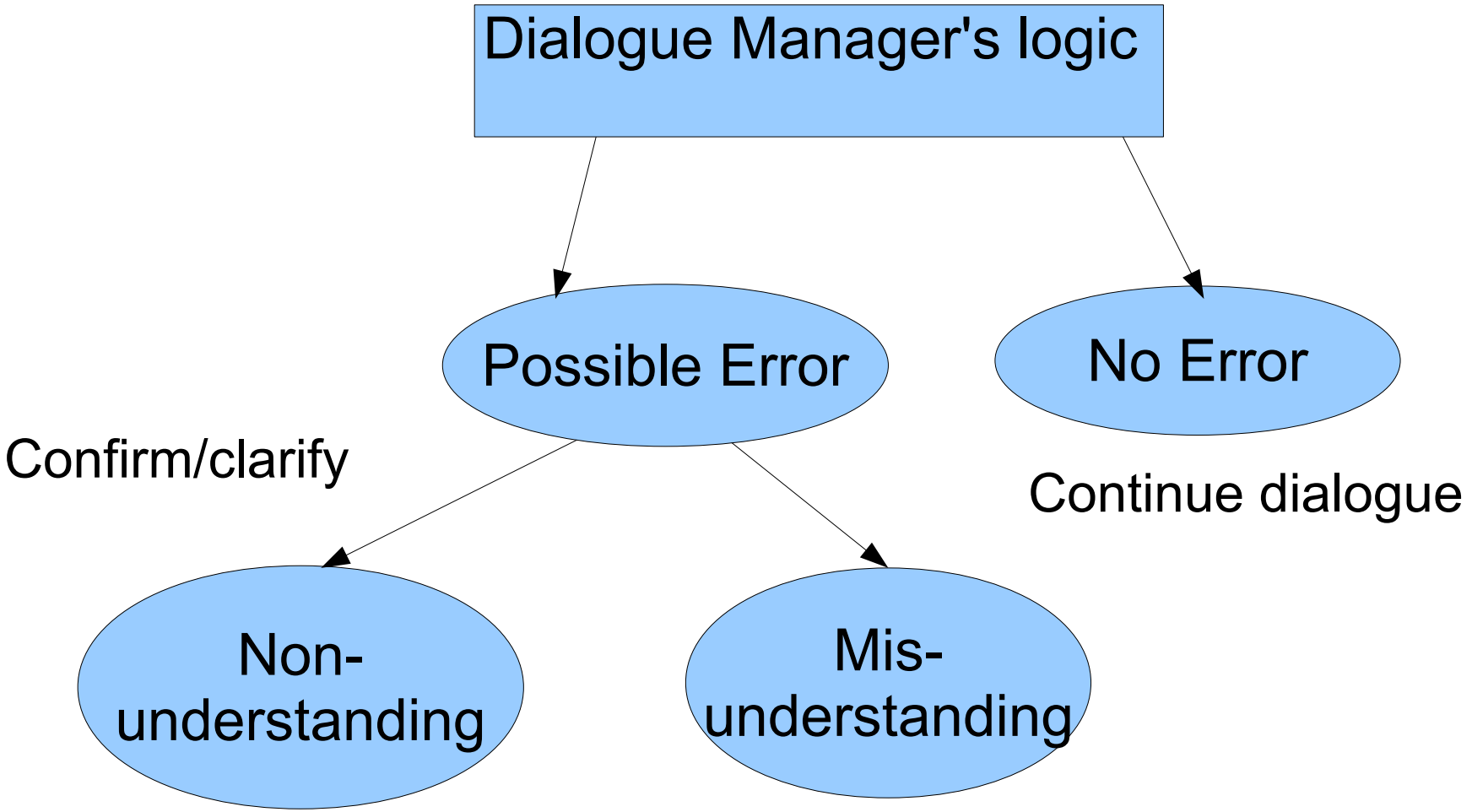
Dialogue Manager's Actions



Confirm/Ask again

Continue dialogue

Dialogue Manager's (DM) Actions



Repeat the question
Ask user to repeat
Help message

Implicit/Explicit confirmation:

Leaving from Waterfront.
Did I get that right?

How does DM Make a Decision?

- Based on features:
  - ASR hypothesis and confidence
    - Confidence is computed from the posterior probabilities
  - Semantic parse (and confidence)
  - Dialogue history
  - Prosodic features help predict if an utterance is misrecognized (Hirschberg, Swertz, Litman, 2004)
    - If user is hyper-articulates, an utterance is less likely to be recognized

Costs of Incorrect Action by DM

- Cost of rejecting hypothesis:
  - A user has to repeat the whole utterance
- Cost of confirming hypothesis:
  - Explicit confirmation elicits simple yes/no answer
  - Implicit confirmation elicits user's response only if recognition was incorrect

Data-driven approaches

- Dialogue Manager Aims to minimize False Rejections and Misunderstandings:

  Bohus and Rudnicky, 2005: Optimizing rejection threshold using supervised machine learning from transcribed data

  - Bohus et al, 2006: Online Supervised Learning of Non-understanding Recovery Policies
    - construct runtime estimates for the likelihood of success of each recovery strategy
    - use these estimates to construct a policy

Data-driven approaches (2)

- Paek & Horvitz 2003
  - Decision making under uncertainty
  - Principle of Maximum Expected Utility:
    - Choose an action so that a utility is maximized
- Skantze 2007
  - a=action h = set of states
  - Estimate utility of action **a** in state **h_i** U(a, h_i)
  - “Choose a grounding action a, so that the sum of all task-related costs and grounding costs is minimised, considering the probability that the recognition hypothesis correct.”

Data Driven Approaches (3)

- Skantze 2007
  - Thresholds are estimated based on dialogue context and state
  - Using efficiency as optimization function
    - Approximation of *user satisfaction* measure
    - Found that efficiency correlates with dialog success
    - Measure "cost" using number of syllables used by both a user and a system

Findings from Skantze 2007

- Different confidence thresholds based on dialogue state

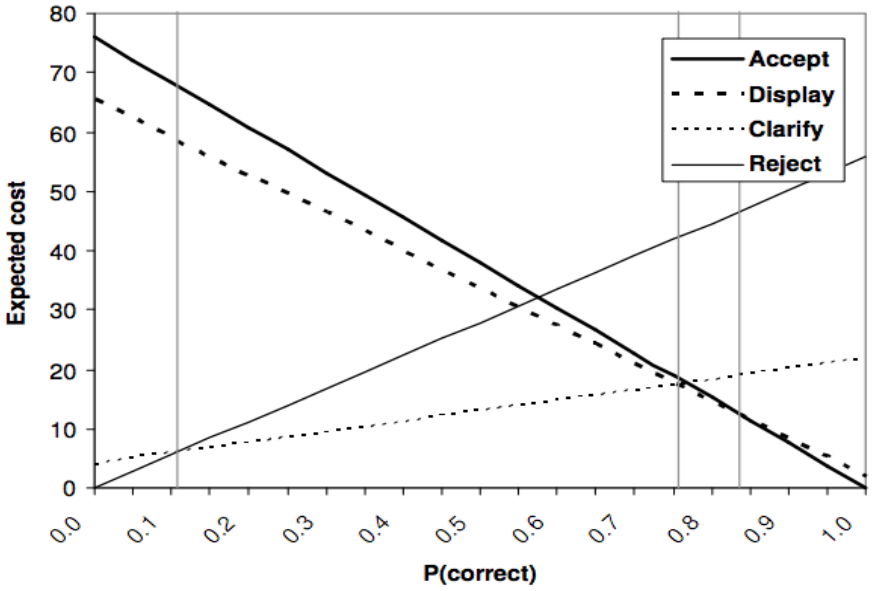


Figure 1: Cost functions and confidence thresholds for grounding the concept MAILBOX after "I can see a mailbox".

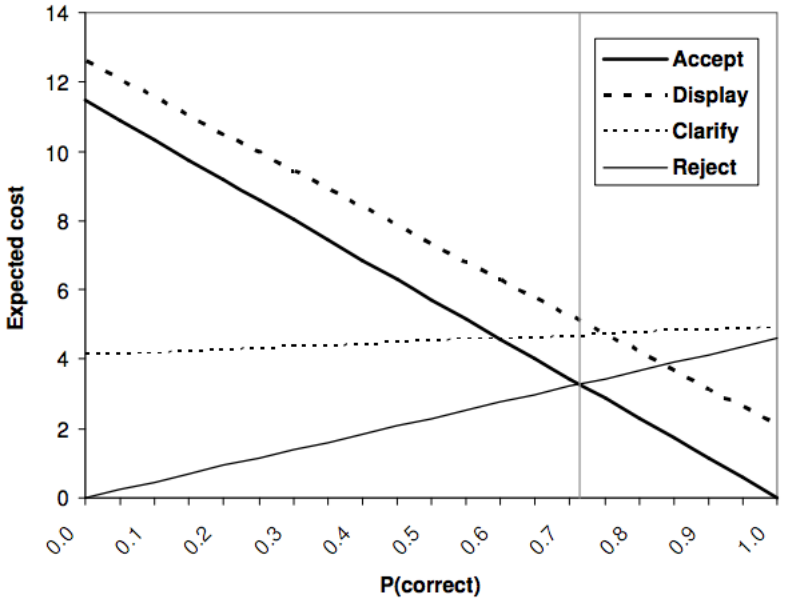


Figure 2: Cost functions and confidence thresholds for grounding the concept TWO after "I can see a two storey building".

Data Driven Approaches (4)

- POMDP Williams & Young (2007)
  - More general and complex approach that handles threshold estimation from data
  - Makes use of parallel recognition hypotheses

Outline

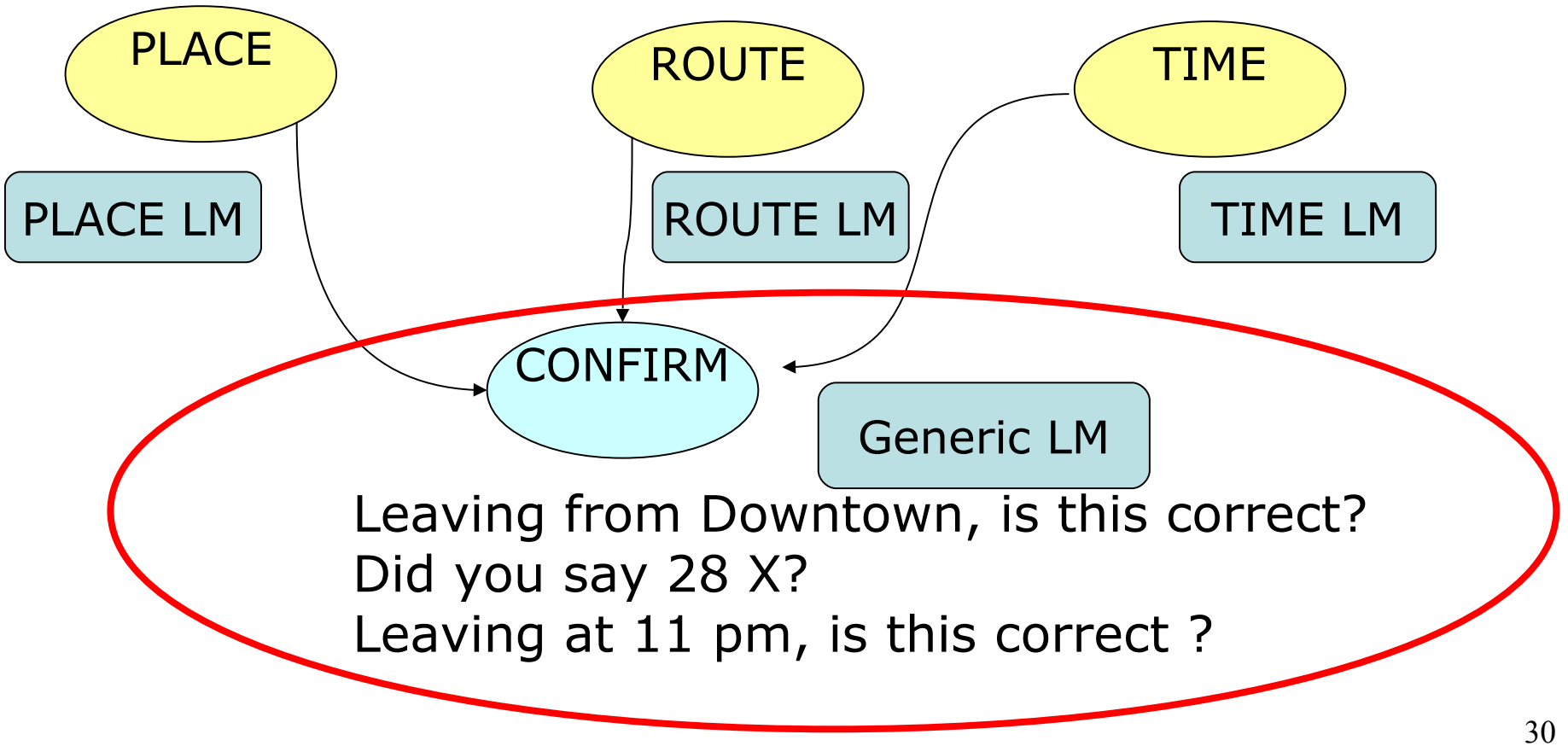
- How do human speakers handle errors in a dialogue
- How do dialogue systems handle errors
  - Overview of dialogue system strategies
- User reactions to system error recovery attempts
- New proposed approach to handling errors

Language Modelling (Let's Go Bus Info System)

Where are you leaving from?

Which route do you need information on?

What time do you want to leave?



How do Users React to Explicit Confirmations?

System's question	User utterance
The **54C** . Did I get that right?	yes you did
Leaving from **ROBINSON**. Is this correct?	**from polish hill**
Going to **WOOD STREET** . Did I get that right?	yes
Going to **REGENT SQUARE** . Is this correct?	**Braddock avenue**
The **61A**. Did I get that right?	wondering when the next bus is

18% user utterances after a confirmation contain a concept

- How can the systems handle this?

Users' Reaction to Errors

- Shin et al. (2002)
  - Annotated error segments: start of an error and back-on-track recovery
  - Airline reservation dialogues
    - (1) SYSTEM tags: explicit confirmation, **implicit confirmation**, help, **system repeat**, reject, non sequitur
    - (2) USER tags: repeat, rephrase, contradict, frustrated, change request, startover, scratch, clarify, acquiesce, hang-up

How Long Does Error Recovery Take?

- 78% of errors recovered (got back on track)
- Average length of error segment for recovered errors is 6.7
- Average length of error segment for unrecovered errors is 10

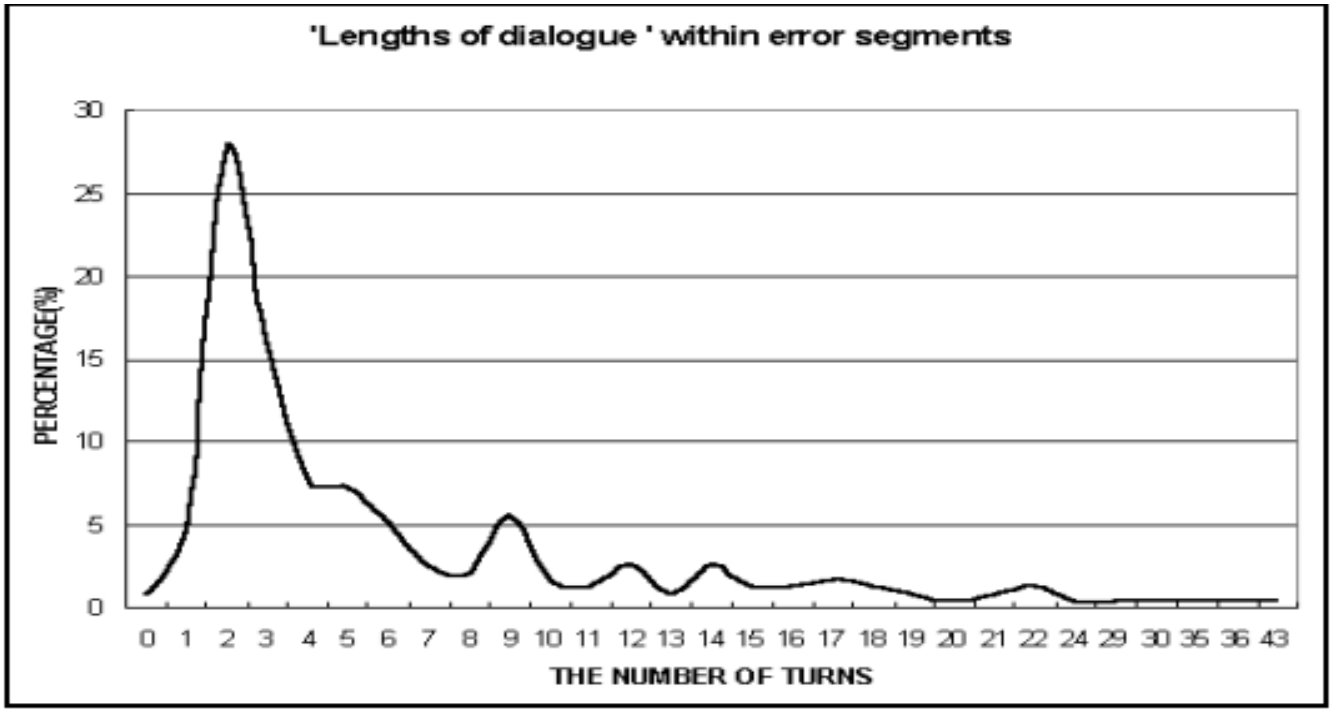


Figure 1: Normalized histogram of the length of error segments (number of turns).

Effects of System actions (Shin 2002)

Error perception	# of err segments	avg err length for BOT	avg err length not BOT	%B OT
Reject	35	6	7.8	83%
Implicit	25	9.6	14.6	68%
Repeat	21	5.8	13	90%
Explicit	10	5.5	8.75	60%

User behaviour after

System repeats
a question:

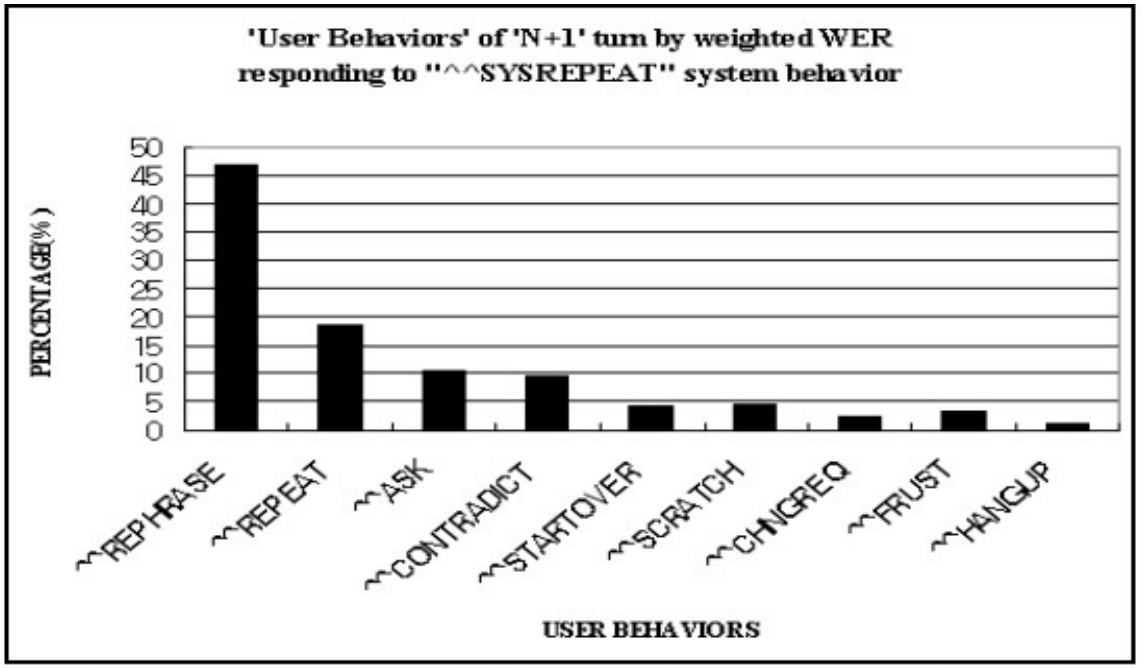


Figure 4: Smoothed Conditional Probability of User Behavior in (N+1)th turn based on weighted WER of 'SYSTEM REPEAT' system behavior in the N-th turn.

System makes an
Implicit Confirmation:

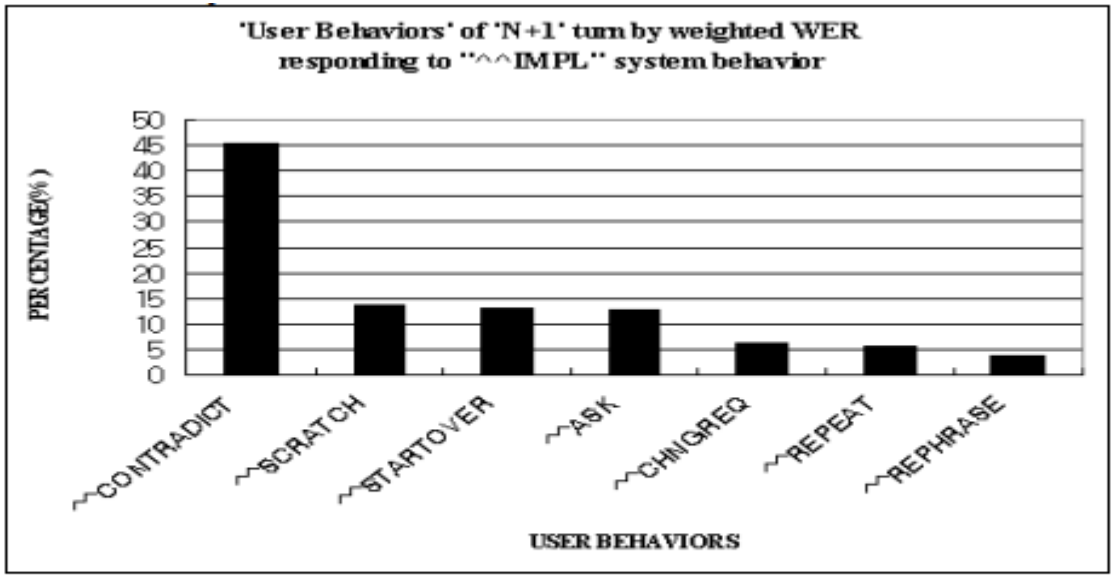


Figure 5: Smoothed Conditional Probability for User Behavior in (N+1)th turn based on weighted WER of 'IMPLICIT CONFIRM' system behavior in the N-th turn..

User Strategies Affect Recovery

- Users in the successful error recoveries
  - use significantly **more rephrasing** than those in the unrecovered errors and **less contradictions**
  - make use of the “start over” and “scratch” features more
  - change travel plans (users are cheating!)

Users Hyperarticulation

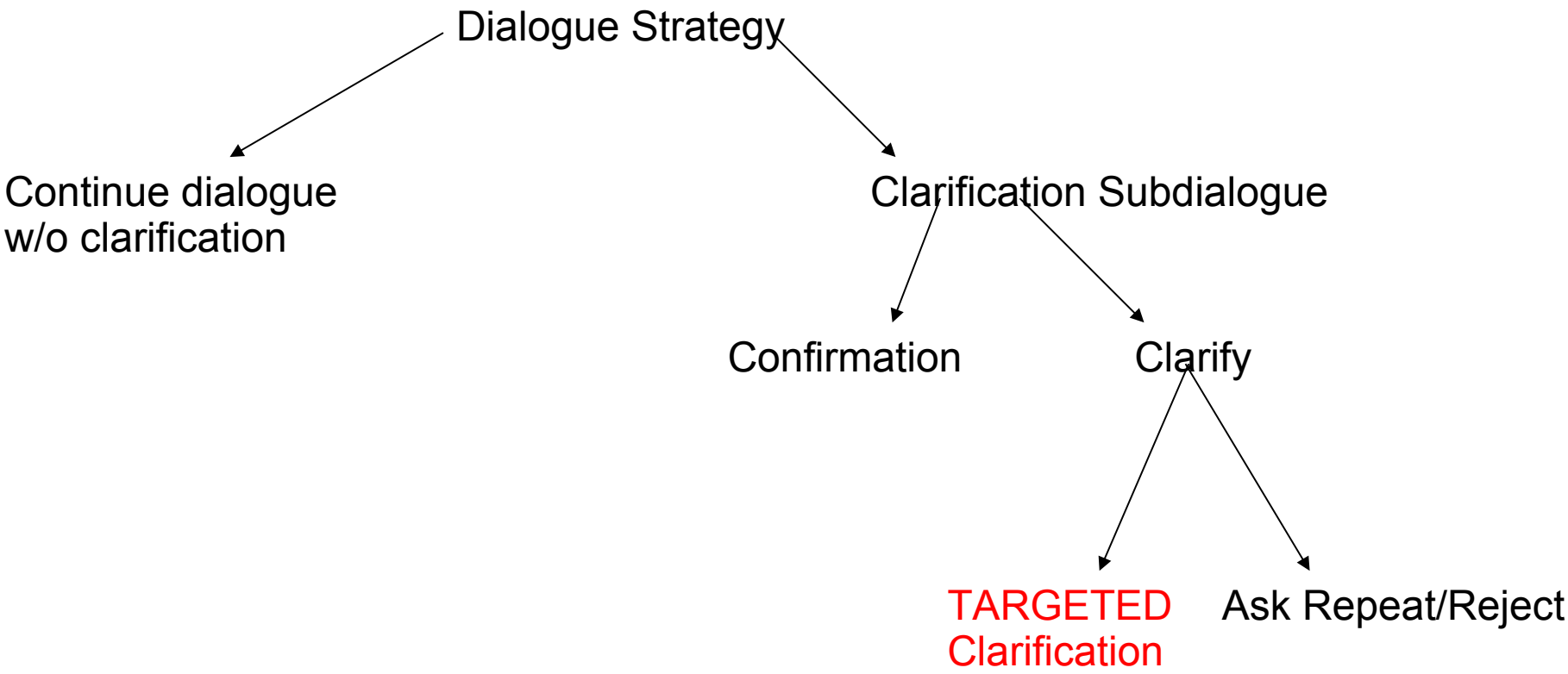
- Swetz et al 2000: compare user utterances that are **corrections** and **non-corrections.** Find that they:
  - differ prosodically, in ways consistent with hyperarticulated speech
- User hyperarticulation is linked to higher ASR error rate
- Users hyperarticulate more after several errors

(play examples)

Outline

- How do human speakers handle errors in a dialogue
- How do dialogue systems handle errors in a dialogue
  - Overview of dialogue system recovery strategies
  - Evaluation of mobile devices
- New proposed approach to handling errors

Dialogue Clarification Strategy



Example with ASR Error

- User: Do you have anything other than these XXX plans?

What clarification Questions would you ask?

Example with ASR error

- User: Do you have anything other than these XXX plans?

Please repeat. Generic ask to repeat

What kind of plans? Targeted 'reprise' clarifications (Purver 2004):

- Ask a directed question about a part of an utterance
- Use recognized words to create a question

Proposed Approach

- Design a system that ask targeted reprise clarification questions.

User: Do you have anything other than these XXX plans?

System: What kind of plans?

Targeted Reprise clarification

[Correct word: floor]

Proposed Approach

User: Do you desire to XXX services to this new clinic?

System: Do I desire to do what?

Targeted Reprise clarification

[Correct words: add new]

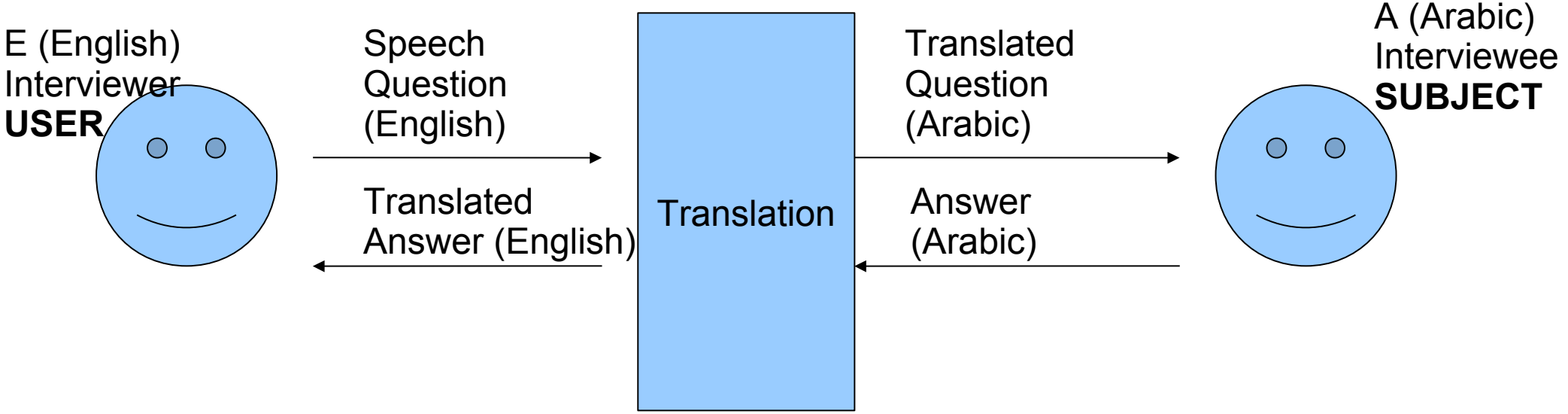
Data: Speech-to-speech Translation System

Domain: English-Arabic Dialogue

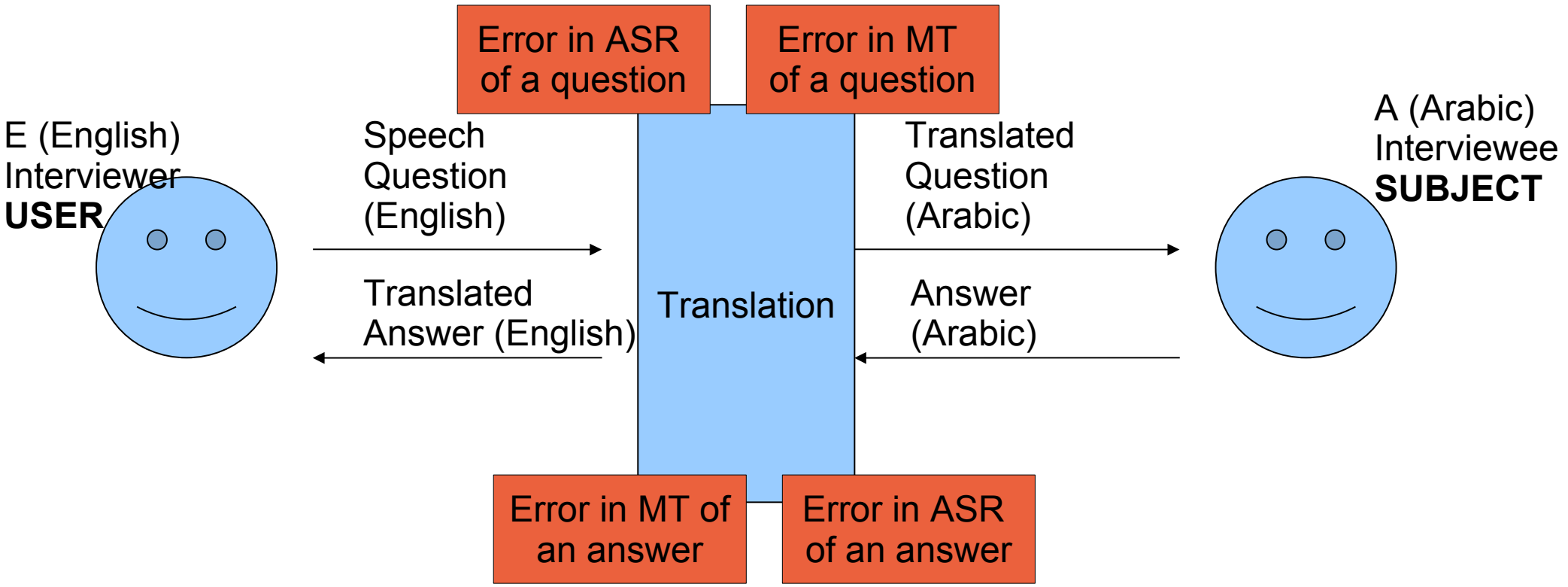
English:	good morning
Arabic:	good morning
English:	may i speak to the head of the household
Arabic:	i'm the owner of the family and i can speak with you
English:	may i speak to you about problems with your utilities
Arabic:	yes i have problems with the utilities

Data collected during evaluation: 3.7K English Utterances

Possibilities of an Error in Speech-to-Speech Translation

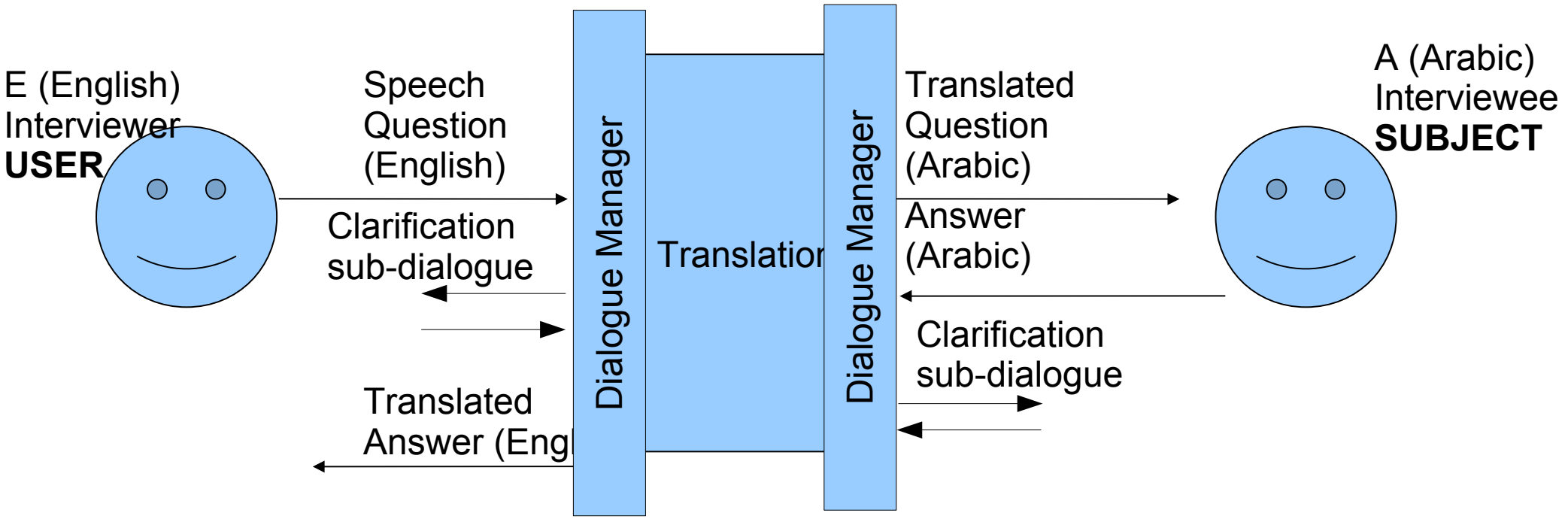


Possibilities of an error in Speech-to-Speech Translation



Ideally these errors are handled by the system

Introducing Clarification Dialogue Manager



Ideally these errors are handled by the system

Dialogue Manager Catches an Error Before it is Translated

E: Good morning my name is major XXX

System: What is your name?

E: major Gareth

Translate to Arabic: “good morning my name is major Gareth”

Requirements

1. Detect a misrecognized segment:

Speech: Do you desire to add new services to this new clinic?

ASR: Do you desire to any services to this new clinic

Requirements

1. Detect a misrecognized segment:

Speech: Do you desire to add new services to this new clinic?

ASR: Do you desire to any services to this new clinic

2. Construct a clarification question from correctly recognized part.

Localized Error Detection

- Use data driven method
- Train a prediction model (Decision tree classifier) to predict if a word is recognized correctly
- Data:
  - 3.7K Utterances (28.6% contain error)
  - 26K Words
- Total words per utt: 7.48
- Misrecognized words in an utt with error 2.03

2-stage approach

- 1. Predict whether an ASR hypothesis contains a recognition error
  - Utterance → correct/incorrect
- 2. For each word in an ASR hypothesis classified as “incorrect” by stage 1, predict if it was recognized correctly
  - Word1 → correct/incorrect
  - Word2 → correct/incorrect
  - …
  - Word N → correct/incorrect

Features: ASR

- Posterior probability generated by speech recognizer from a user's utterance, acoustic, language models
  - 1. whole utterance
  - 2. in current word; average over 3 words; whole utterance

Features: Prosodic

- Features extracted from speech signal
  - F0(MAX/MIN/MEAN/STDEV)
  - energy(MAX/MIN/MEAN/STDEV)
  - proportion of voiced segments
  - duration
  - timestamp of beginning of first word
  - speech rate

Features: Syntactic

- Part-of-Speech Tags on a hypothesis

UTT: hello my name is sergeant inman

ASR: hello my name is sergeant in in

ASR POS tags: hello/UH my/PRP name/NN is/VBZ

sergeant/NN in/IN in/IN

Features: Syntactic

ASR POS tags: hello/UH my/PRP name/NN is/VBZ sergeant/NN in/IN in/IN

Stage1 (Utterance): count of unigrams and bigrams:

UH (1); PRP (1); NN (2); VBZ(1); IN (2)

UH_PRP (1) ; PRP_NN (1); NN_VBZ (1); VBZ_NN (1); NN_IN (1); IN_IN (1)

Features: Syntactic

ASR POS tags: hello/UH my/PRP name/NN is/VBZ sergeant/NN in/IN in/IN

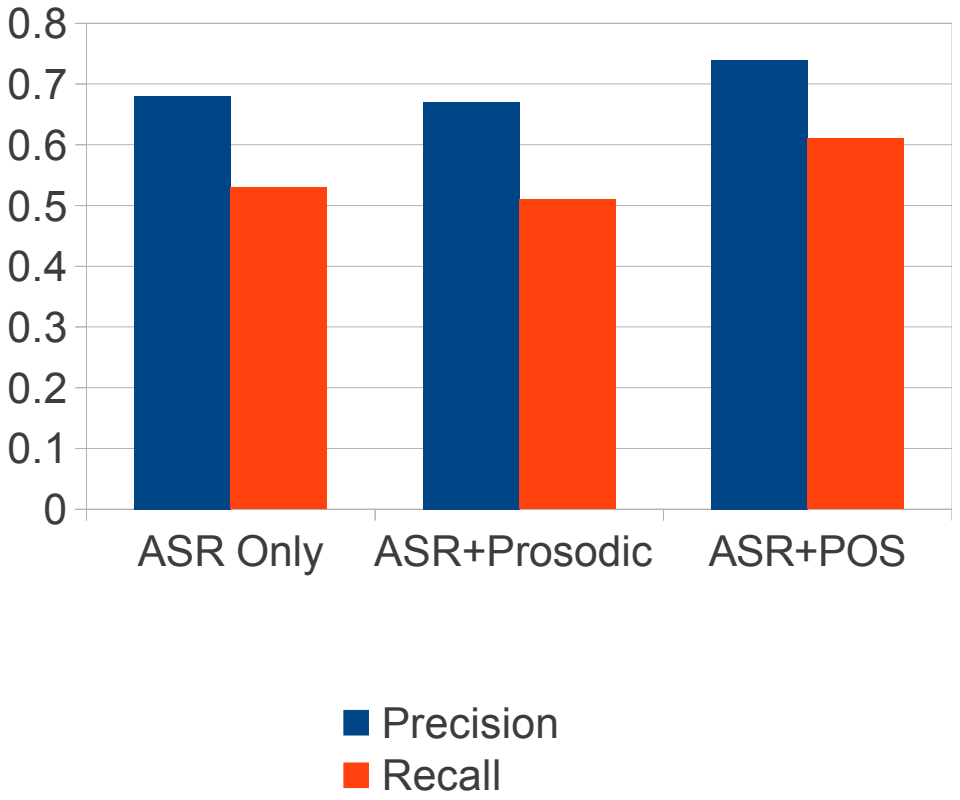
Stage2 (Word): POS tag of this word, previous word, next word

POS	this	prev	next
hello:	UH,	- ,	PRP
my:	PRP,	UH,	NN
name:	NN,	PRP,	VBZ
...			
in:	IN,	NN,	IN

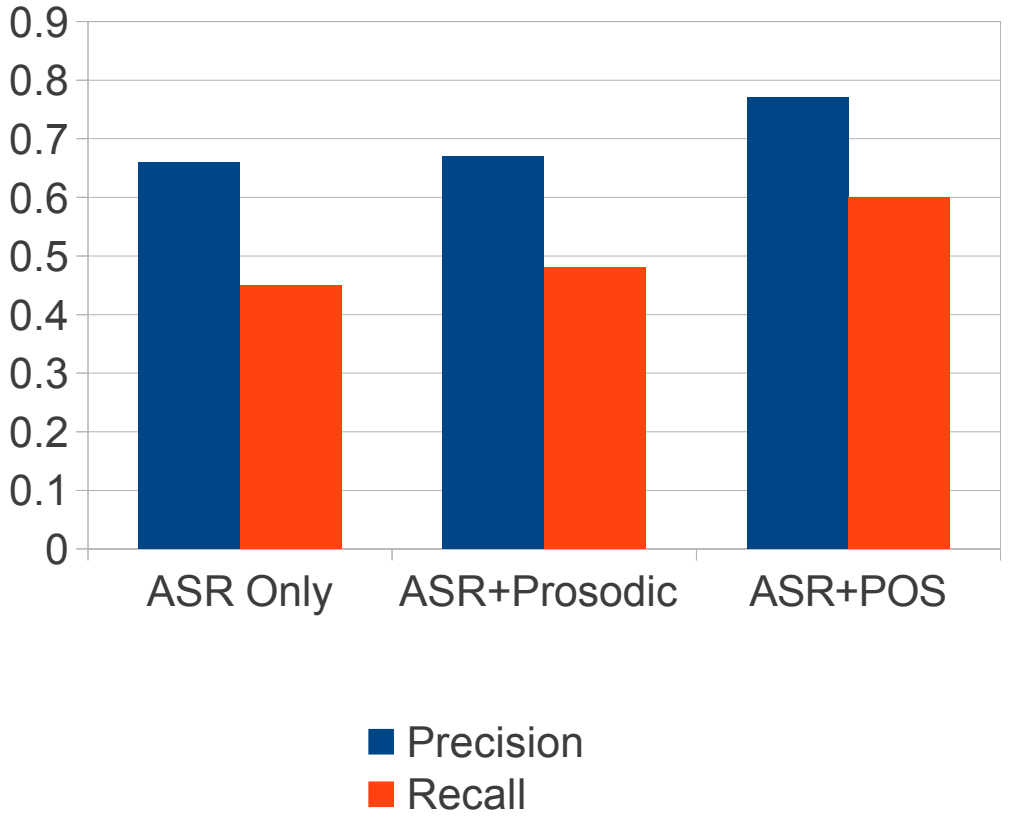
Method

- Machine learning on transcribed data
- Using WEKA machine learning tool
- Results from 10-fold cross-validation
  - Precision/recall of detecting incorrect utterance/word
  - Majority baseline – always predict “correct” (recall of predicting incorrect = 0)

Evaluation Results: Predict Utterance Recognition



Evaluation Results: Predict Word Recognition



Clarification Dialogue Modelling

What type of question should the system ask?

How to construct a question?

Mechanical Turk Experiment

Show a sentences with misrecognized word removed
e.g. "XXX these supplies stolen"
Ask users:

- Can XXX be omitted without change of meaning? (yes/no)
- Can you guess the missing word(s)?
- Can you guess a POS of the missing word(s)?
- Can you ask a clarification question?
- What type of question is it (reprise clarification/confirmation/general repeat)?

Preliminary Results

- 25 sentences x 3 annotators
- Correctly guessed words:

Preliminary Results

- 25 sentences x 3 annotators
- Correctly guessed words: 16-20%
- Correctly guessed POS tags:

Clarification Questions: Preliminary Results

- 25 sentences x 3 annotators
- Correctly guessed words: 16-20%
- Correctly guessed POS tags: 36-52%

Clarification Questions: Preliminary Results

- Users asked different questions:
  - Example ASR output:
    - what kind of jewellery was XXX
  - Questions generated by 3 different users:
    - What did you want to know about the jewellery?
    - What was the last word you said?
    - Would you repeat that?

How will we use this data?

- Collect data for 700 sentences (~ 6 months)
- Learn dialogue strategy (dialogue system's action) from the data
- Features:
  - POS tag of error word(s)
  - Position of error word(s)
  - Semantic role of the error segment (subject/object)

Clarification Questions: Preliminary Results

Out of 44 questions:

- Reprise clarification: 31
- Ask to repeat: 10
- Confirmation: 3

Summary

- How do human speakers handle errors in a dialogue
  - Using diverse strategies
- How do dialogue systems handle errors
  - System's Actions: Repeat Question, Ask User to Repeat/Rephrase, Explicit/Implicit Confirmation, Play Help Message
  - Systems try to choose best possible action to get conversation back-on-track
  - Use rule-based or machine learning approaches
- How do users react to system errors

End Note

- You can not foresee all possible user actions

Thank you

Questions?

sstoyanchev@cs.columbia.edu

References

- *M. Purver* The Theory and Use of Claricaton Requests in Dialogue, PhD thesis, 2004
- *D. Bohus and A. Rudnicky*. A principled approach for rejection threshold optimization in spoken dialog systems. In INTERSPEECH 2005.
- *Bohus, D. et al.* Online Supervised Learning of Non-understanding Recovery Policies, in SLT-2006
- *Julia Hirschberg, Diane J. Litman, and Marc Swerts*. Prosodic and other cues to speech