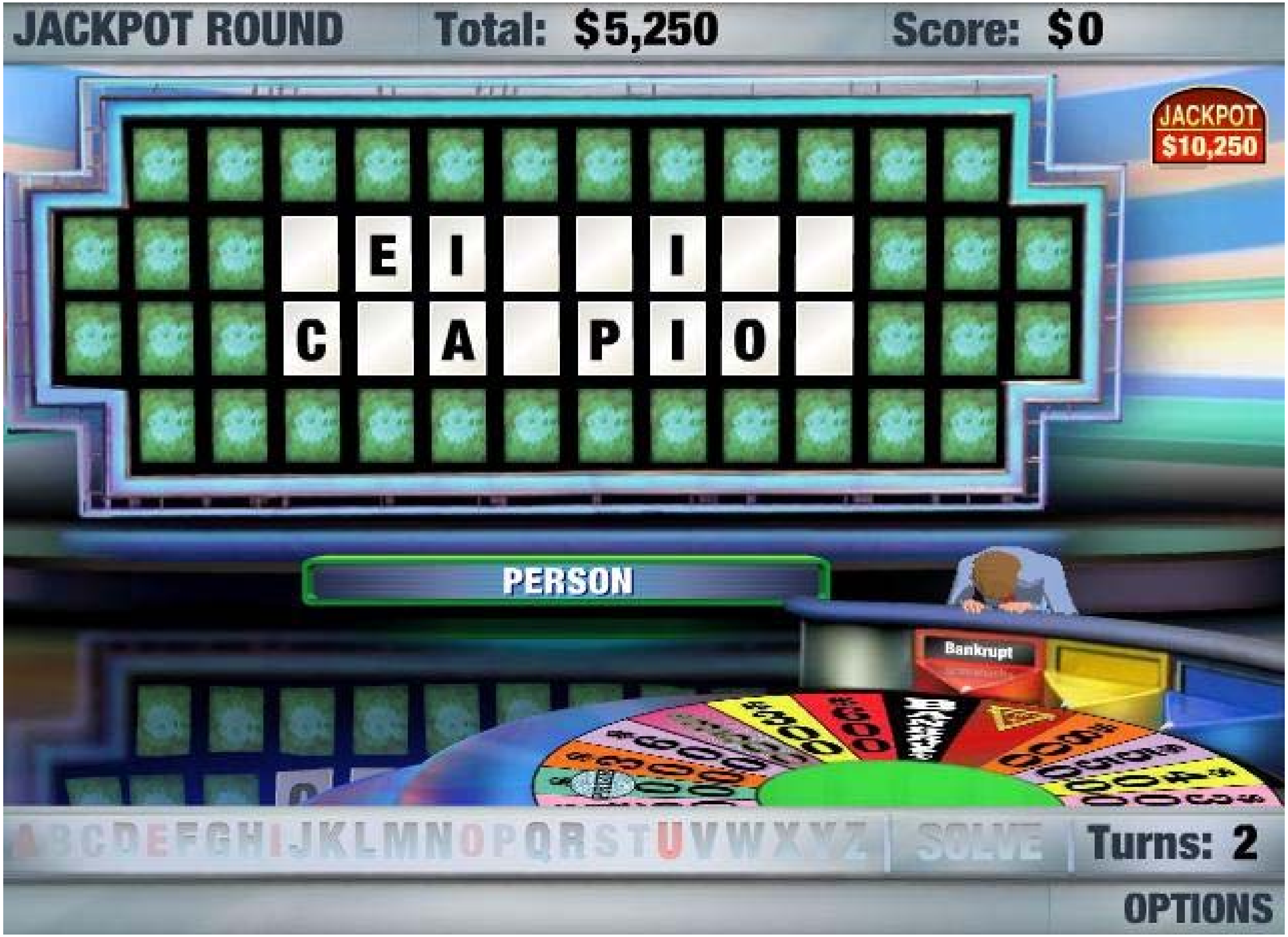
Help Me Understand You:
Addressing the Speech Recognition Bottleneck

Rebecca Passonneau,* Susan Epstein[†] and Joshua Gordon*

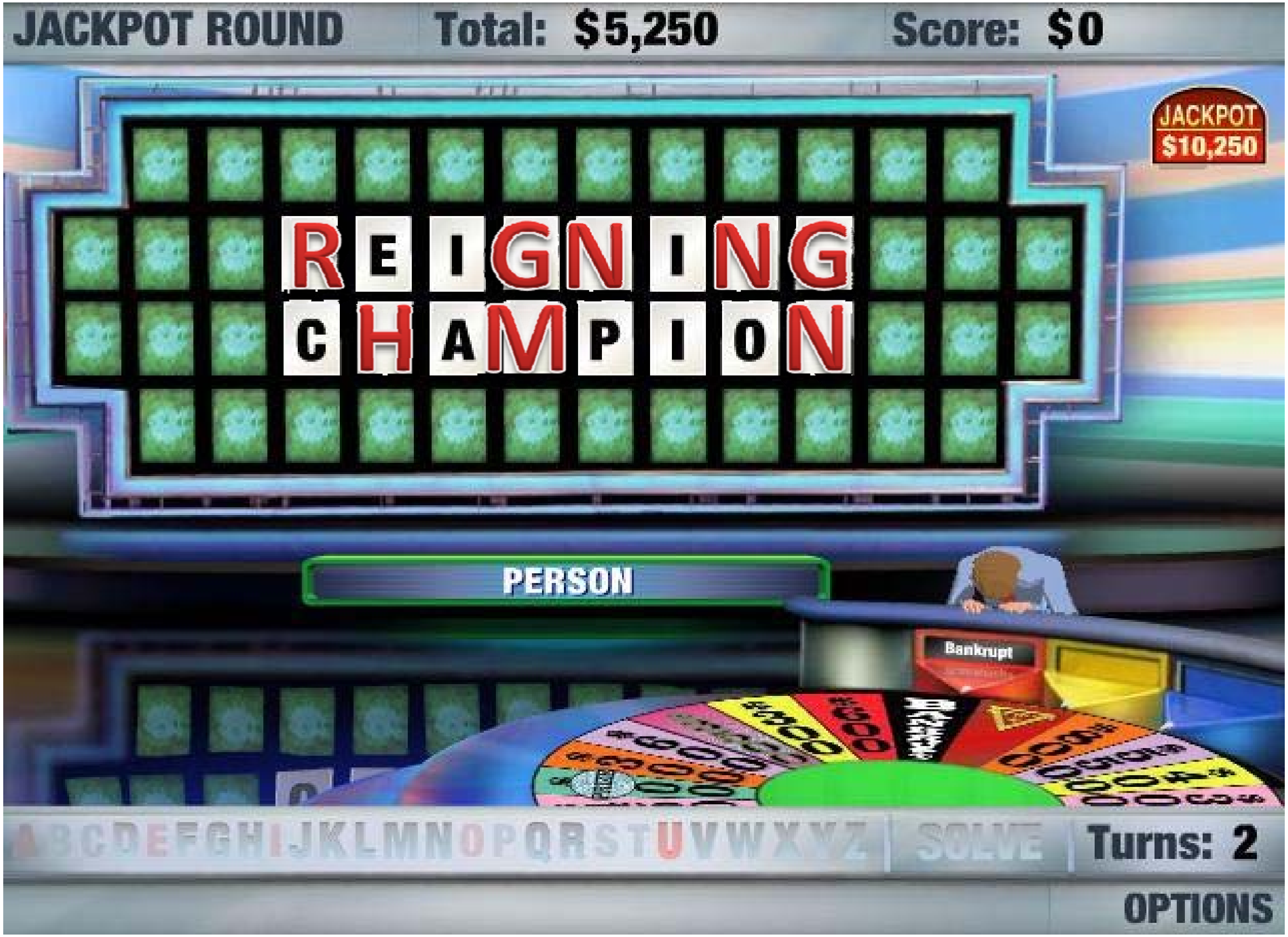
*Columbia University

[†] Hunter College and The Graduate Center of The City University of New York

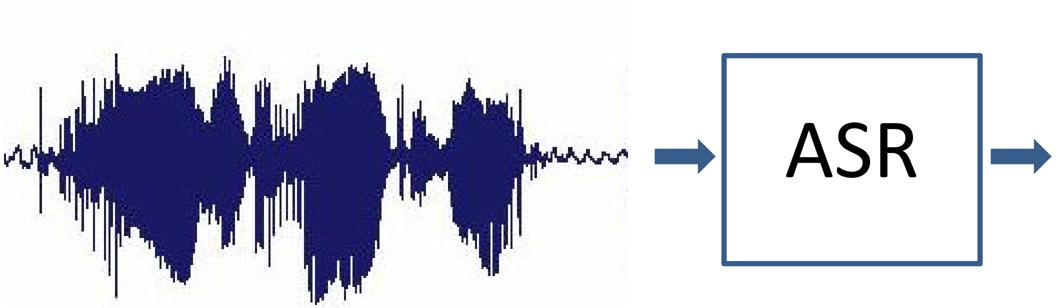
Jeopardy: Text through a *Noisy Channel*



Domain Knowledge Helps: *PERSON* + C_A_MPIO_



Automatic Speech Recognition (ASR): A Noisy Channel



ASR

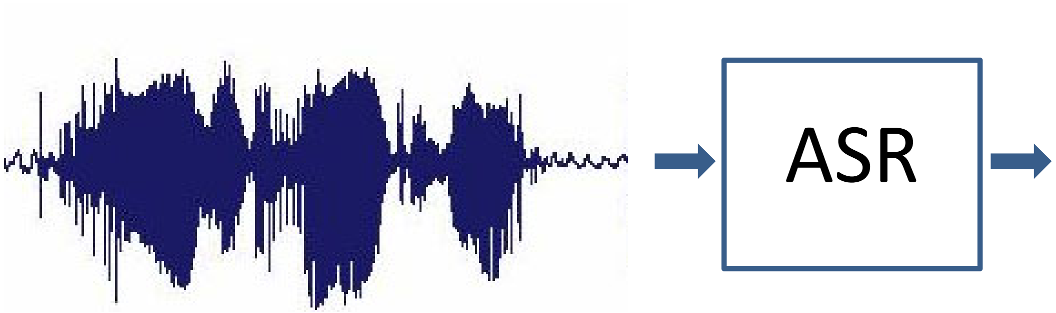
Great Expectations

Grape Expectations

Great Expeditions

ASR for Book Titles	**Google Books**
INTO THAN NINE	*TO THE NINES*
OF 5 PEOPLE UNION HEAVEN	*THE 5 PEOPLE YOU MEET IN HEAVEN*
WHAT HEART INTO	*POUR YOUR HEART INTO IT*
ABORT BANDIT	*BALD BANDIT*
SWEET NINE STORIES	*SWEET LAND STORIES*
HUMOROUS REMEMBER THIS	*HUMOROUS TEXTS*
ELUSIVE TOTAL NAH	*?*
DOING SORROW RUN	*?*
PEOPLE EXIT	*?*
ROLL DWELL	*?*

Automatic Speech Recognition (ASR): A Noisy Channel



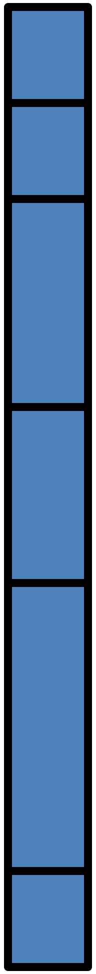
Great Expectations

Grape Expectations

Great Expeditions

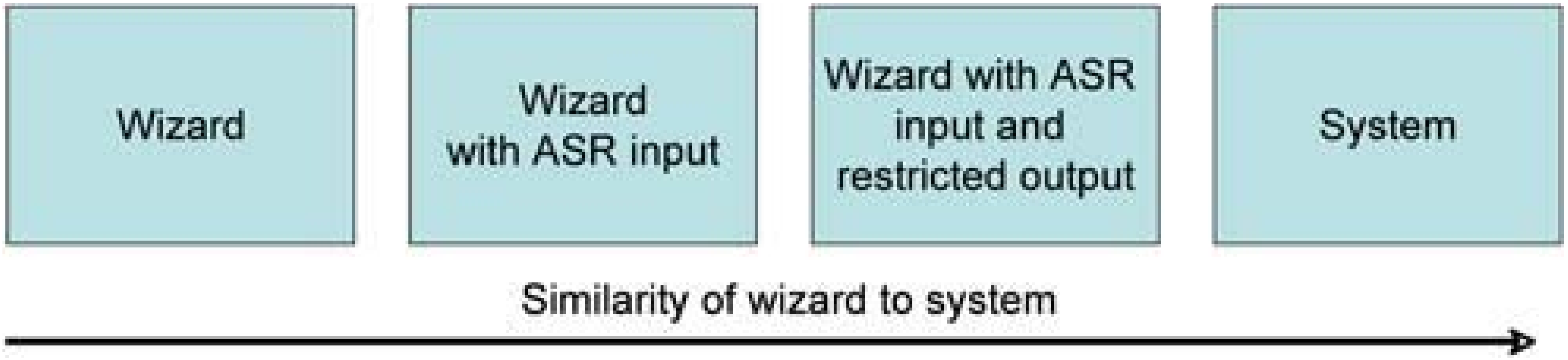
ASR for Book Titles	**Google Books 2/10**
INTO THAN NINE	*TO THE NINES*
OF 5 PEOPLE UNION HEAVEN	***THE 5 PEOPLE YOU MEET IN HEAVEN***
WHAT HEART INTO	*POUR YOUR HEART INTO IT*
ABORT BANDIT	*BALD BANDIT*
SWEET NINE STORIES	***SWEET LAND STORIES***
HUMOROUS REMEMBER THIS	*HUMOROUS TEXTS*
ELUSIVE TOTAL NAH	*?*
DOING SORROW RUN	*?*
PEOPLE EXIT	*?*
ROLL DWELL	*?*

Outline



- Wizard ablation
- CheckItOut dialogue system and application domain
- Pilot experiment: book title recognition
    - ASR noisy channel
    - Domain knowledge of book titles
- Results
    - Correct title found 70%
- Discussion
    - Previous work: sometimes erroneous ASR best ignored
    - Our pilot: erroroneous recognition useful for retrieval
- Current and future work

Loqui Dialogue Project: Wizard Ablation


Wizard
Wizard with ASR input
Wizard with ASR input and restricted output
System
Similarity of wizard to system

- Adapt conventional Wizard of Oz (WOz) paradigm
  - Ideal human-machine dialogue will differ from human-human dialogue
  - Ablated wizards apply human intelligence to component technologies
- Collect corpora (sets of dialogues) that vary in degree of ablation
- Evaluate dialogues across conditions (PARADISE, Walker et al 1997)
  - For task success
  - For user satisfaction
- Apply machine learning to distinct corpora
  - Learn what ablated wizards do
  - Determine which corpora are the best "teachers"

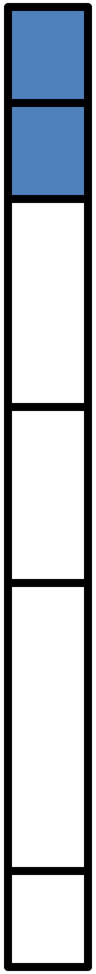
Related Work

- Learning dialogue strategies from corpora
  - Initial work in early 2000s (Levin, Pieraccini & Eckert, 2000; Scheffler & Young 2002)
  - Has become the dominant approach for dialogue management
- WOz with ASR input to wizards
  - Zollo 1999
  - Skantze 2003
- Other alternatives to human-human corpora
  - Simulated dialogue corpora (Schatzmann et al. 2005; Ai & Litman 2006)
  - WOz + simulation (Griol et al., 2008)

CheckItOut Domain: Library Transactions

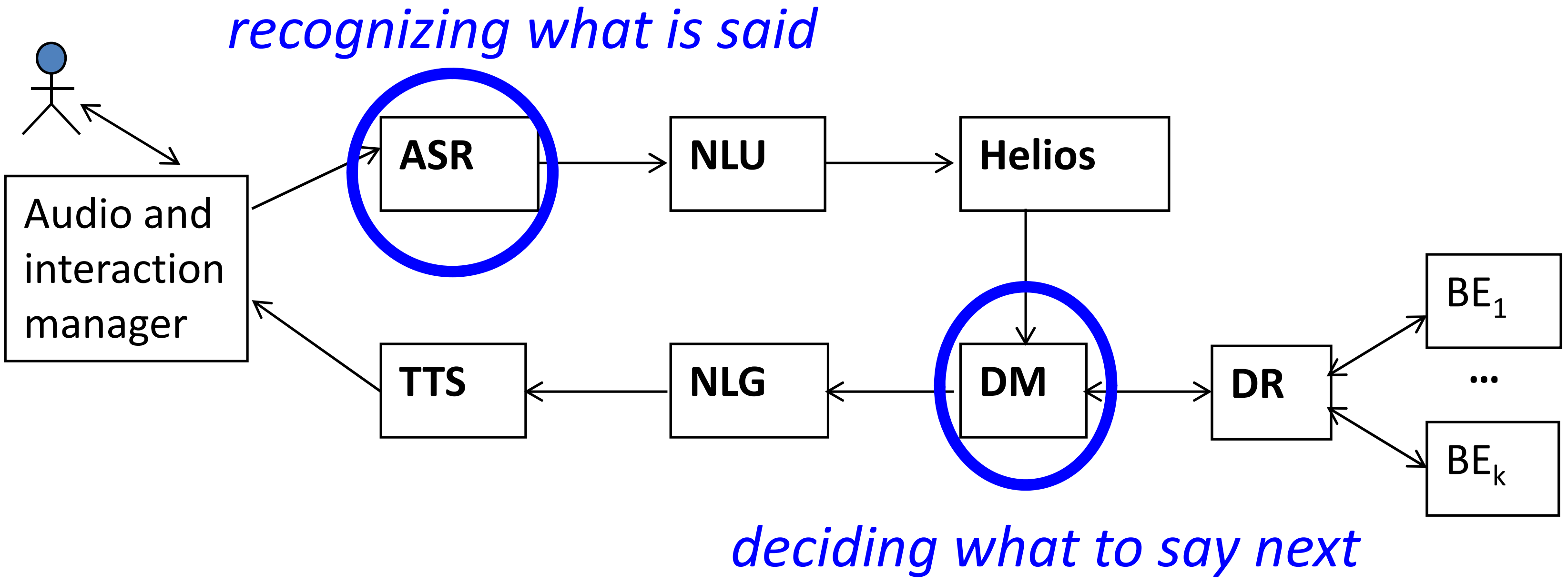
- Andrew Heiskell Braille and Talking Book Library
    - Branch of New York City Public Library
    - Branch of National Library Service
- Book transactions
    - Callers order books/cassettes by telephone
    - Orders sent/returned by U.S.P.O.
- CheckItOut database (Postgres)
    - Replica of Heiskell Library book catalogue (N=71,166)
    - Mockup of patron database for 5,028 currently active patrons
- CheckItOut Dialog Model
    - Based on Loqui Human-Human Corpus (175 recorded calls)
    - Domain independent error handling and repair
    - Domain dependent task hierarchy to guide the dialogue manager

Loqui Human-Human Corpus: Sample Book Request



Caller: I don’t think she had this <pause> particular book uh Jasons Yukon Gold
Caller: She was wondering if you have that
Caller: She read the sequel just now
Librarian: Okay
. . .
Librarian: the title is Jasons Yukon [Gold]
Caller: [I] think so I have a number here
Caller: I think it's RC <pause> one two seven eight six
Caller: Is that right
Librarian: mmm that's Tender Mercies
Caller: okay how about this five zero two o one
Caller: and I have a bunch of numbers here
Librarian: Jasons Gold right
Caller: oh Ji- Ja- Jasons Gold [then]
Librarian: [yeah]
Caller: yeah could you uh send that when y- if you have it <pause> t- to her

CheckItOut Dialogue System



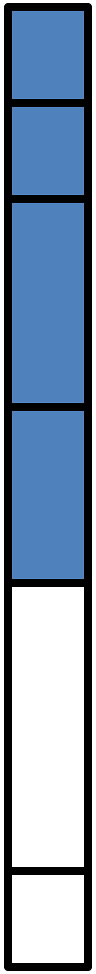
Carnegie Mellon University's Olympus/Ravenclaw

- ASR: Automatic Speech Recognition
- NLU: Natural Language Understanding
- Helios: Confidence Annotation
- DM: Ravenclaw dialog manager
- DR: Domain Reasoner
- NLG: Natural Language Generation
- TTS: Text-to-speech synthesis

Pilot Study: Offline Wizards Interpret ASR for Booktitles

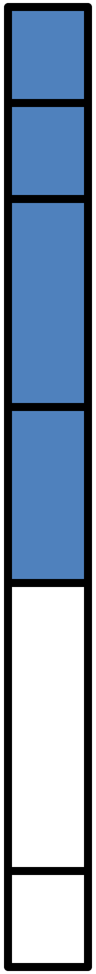
- Participants
  - Callers: two undergraduates at Hunter College (A, B), one researcher (D)
  - Offline wizards: three Hunter undergraduates (A, B, C)
- Recognizer data
  - Dictionary of words based on 500 titles (1400 words)
  - Unigram frequencies (individual words, no bigrams, trigrams)
- Materials
  - Three disjoint sets of 50 titles
  - Each caller produced ASR for one set of titles
  - Each wizard received ASR for one title set (wizard ≠ caller)
  - Each wizard received a text file of the full title list (N=71,166)
- ASR performance in Word Error Rate (WER)
  - D: 0.69
  - A: 0.75
  - B: 0.83
- Task
  - For each ASR string, find the most likely title
  - Document their thoughts

Moderately Difficult Examples



INTO THAN NINE	TO THE NINES INTO THE INFERNO INTO THE NIGHT INTO THE WILD
OF 5 PEOPLE UNION HEAVEN	THE 5 PEOPLE YOU MEET IN HEAVEN NO TELEPHONE TO HEAVEN A LONG WAY FROM HEAVEN DO THEY WEAR HIGH HEELS IN HEAVEN
ROLL DWELL	CROMWELL ROBERT LOWELL ROAD TO WELLVILLE ROAD TO WEALTH

Difficult Examples



WHAT HEART INTO

WHAT THE HEART KNOWS
THE LAST INHERITOR
A PRIVATE VIEW

ELUSIVE TOTAL NAH

LUSITANIA
THE ELUSIVE FLAME
I LIVED TO TELL IT ALL

PEOPLE EXIT

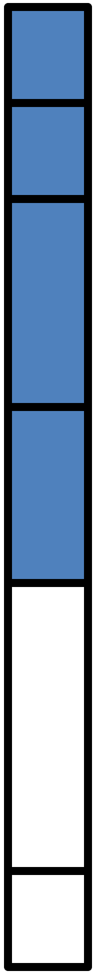
PEOPLE IN TROUBLE
PEOPLE VERSUS KIRK
THE ODES OF PINDAR

Results

	Wizard A		Wizard B		Wizard C	
Category	**Count**	**%**	**Count**	**%**	**Count**	**%**
Correct	30	**66.7**	33	**71.7**	33	**71.7**
Ambiguous	0	0.0	4	8.7	0	0.0
Incorrect	7	15.5	1	2.2	13	28.3
No response	8	17.8	8	17.4	0	0.0
Total	45	100.0	46	100.0	46	100.0

- Wizards are correct 70% of the time on average
- Wizards behaved differently when uncertain
  - A: about evenly divided between “Incorrect” and “No response”
  - B: same proportion of “No response” as A; identified “Ambiguous” cases
  - C: always responded -- higher proportion of “Incorrect”

Strategies

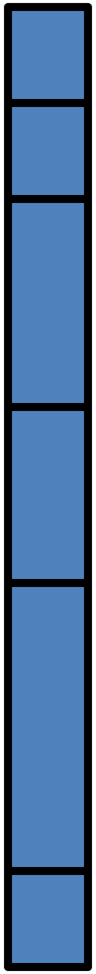


	A		B		C	
	#	%	#	%	#	%
Word hits	11	**24**	17	**37**	13	**28**
Lexical Rarity	5	11	3	7	0	0
Word hits +location	2	4	3	7	13	**28**
Word hits +lexical rarity	1	2	5	11	2	4
Word hits +lexical rarity + location	11	**24**	5	11	0	0
Phonetic	8	18	6	13	1	2
Semantic	1	2	1	2	0	0
Other	6	13	6	13	17	37
Totals	45	99	46	100	45	100

Discussion

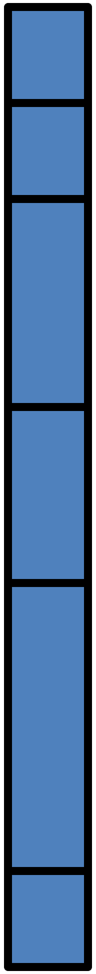
- Previous work: erroneous ASR can sometimes be ignored
    - Zollo 1999, evacuation plan
        - 7 WOz dialogues
        - WER=0.30
        - Wizards signaled a misunderstanding only 35% of the time that the ASR was incorrect
    - Skantze 2003, navigation task
        - 40 dialogues (5 scenarios per 8 distinct wizard/user pairs)
        - WER=0.42
        - Wizards rarely signaled misunderstanding (5% overall)
        - Wizards responded to non-understanding (20% overall) by continuing a route description, asking a task related question, or asking for clarification
- Erroneous ASR should be incorporated into backend retrieval, cf:
    - Machine Translation + Information Retrieval
    - Voice search, e.g., mobile devices with access to web
    - String matching with errors (edit distance; soundex)

Current Work



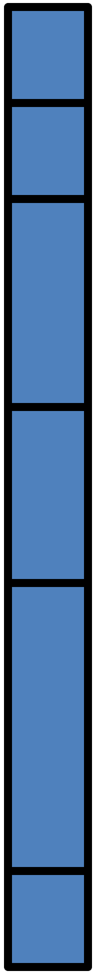
- Online version of same experiment, 4200 data points
  - 7 participants, alternating as wizard/caller (21 * 2 distinct pairs)
  - 5 sessions per participant
  - 20 titles per session
  - Realistic language model (7500 words, bigram model)
  - WER=0.71
  - Backend query function using string matching with errors
- Ratcliff/Obershelp string matching
  - |Matching characters|/|Total characters|
  - Matching characters = recursively find longest common subsequence of 2 or more characters

Moderately Difficult Examples with Ratcliff/Obershelp



INTO THAN NINE	TO THE NINES	0.74
	INTO THE INFERNO	0.73
	INTO THE NIGHT	0.70
	INTO THE WILD	0.59
OF 5 PEOPLE UNION HEAVEN	**THE 5 PEOPLE YOU MEET IN HEAVEN**	0.72
	NO TELEPHONE TO HEAVEN	0.57
	A LONG WAY FROM HEAVEN	0.50
	DO THEY WEAR HIGH HEELS IN HEAVEN	0.45
ROLL DWELL	CROMWELL	0.67
	ROBERT LOWELL	0.61
	ROAD TO WELLVILLE	0.52
	ROAD TO WEALTH	0.50

Difficult Examples with Ratcliff/Obserhelp



WHAT HEART INTO	WHAT THE HEART KNOWS	0.74
	THE LAST INHERITOR	0.61
	I CAN'T FORGET YOU	0.42
	A PRIVATE VIEW	NA
ELUSIVE TOTAL NAH	LUSITANIA	0.62
	THE ELUSIVE FLAME	0.57
	I LIVED TO TELL IT ALL	0.56
PEOPLE EXIT	PEOPLE IN TROUBLE	0.64
	PEOPLE VERSUS KIRK	0.62
	THE ODES OF PINDAR	NA

Future Work

Book title requests in context of full dialogue

- Recognize a “title request” utterance (examples below)
- Semantic interpretation of the utterance
    - Classification of utterance type (e.g., title request)
    - Integrate with backend query

Examples from Transcripts		
“Front Matter” of Title Utterance	**Title Utterance**	**Actual Title**
but it’s	prince of beverly hills	The Prince of Beverly Hills
we were wondering if you had	evidence that demands a verdict	Evidence that Demands a Verdict
what is the	the next uh uh installment	Remembrance of Things Past: Volume II
I’d like to try um	age of innocence	The Age of Innocence