

# Dialog in NLP applications

---

VELJKO MILJANIC



# Overview

---

## Applications in S2S systems

- Overview of S2S system architecture
- Modeling contextual information in S2S
- Improving S2S systems with DA tags and word prosodic prominence
- Transonics S2S system

## Applications in web search

- Using dialog systems to improve voice search
- Using web search data to improve dialog systems

# Speech to speech

---

Spoken phrases are instantly translated and spoken in a second languages

- Skype translator

Typically realized as three independent tasks

- Source speech transcription (ASR)
- Translation of source text to target text (MT)
- Synthesizing target speech (TTS)



# S2S with contextual information

---

Enriching machine-mediated speech-to-speech translation using contextual information

- Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore and Shrikanth Narayanan

Contextual information benefits

- Augment the output hypothesis to improve understanding and disambiguation
- Improve machine translation
- Improve quality of text-to-speech
- Aid in the natural flow of the dialog

# Adding Contextual Information to S2S Model

---

$$S_t^* = \arg \max_{S_t} P(S_t|S_s) \quad (1)$$

$$P(S_t|S_s) = \sum_{T_t, T_s, L_s, L_t} P(S_t, T_t, T_s, L_s, L_t|S_s)$$

$$= \sum_{T_t, T_s, L_s, L_t} P(S_t|T_t, L_t, T_s, L_s, S_s) \cdot P(T_t, L_t|T_s, L_s, S_s) \cdot P(L_s|T_s, S_s) \cdot P(T_s|S_s) \quad (2)$$

$$\approx \sum_{T_t, T_s, L_s, L_t} P(S_t|T_t, L_t, L_s) \cdot P(T_t, L_t|T_s, L_s) \cdot P(L_s|T_s, S_s) \cdot P(T_s|S_s) \quad (3)$$

$$\max_{S_t} P(S_t|S_s) \approx \max_{S_t} P(S_t|T_t^*, L_t^*, L_s^*) \cdot \max_{T_t, L_t} P(T_t, L_t|T_s^*, L_s^*) \cdot \max_{L_s} P(L_s|T_s^*, S_s) \cdot \max_{T_s} P(T_s|S_s) \quad (4)$$

**Augmented**

**Text-to-Speech**

**Enriched**

**Machine Translation**

**Rich Annotation**

**Speech Recognition**




# Extracting Contextual Information

---

## Dialog act tags

- Maxent classifier is use to estimate DA conditional probability
- Lexical, syntactic and acoustic features within a bounded local context
- Trained on Switchboard-DAMSK corpus
  - Accuracy 70.4% on 42 tags and 82.9% on 7 tags
  - statement, acknowledgment, abandoned, agreement, question, appreciation and other

## Prosodic word prominence

- 4.7h Switchboard audio hand-labeled for pitch accent markers
  - Pitch markers are mapped to words as two classes: accent and none
  - 78.5% accuracy
- 

# Source enrichment: phrase-based translation

---

## Phrase based translation

- Phrase translation table: probabilities of phrase translation pairs
- Target language model: probability of output word sequence

Contextual information is added by conditioning phrase translation table and language model on it:

$$\begin{aligned} T_t^* &= \arg \max_{T_t} P(T_t | T_s, L_s) \\ &= \arg \max_{T_t} \frac{P(T_s | T_t, L_s) \cdot P(T_t | L_s)}{P(T_s | L_s)} \\ &= \arg \max_{T_t} P(T_s | T_t, L_s) \cdot P(T_t | L_s) \end{aligned}$$

# Source enrichment: phrase-based translation

---

Conditioning on contextual information is increasing number entries in phrase table and language model

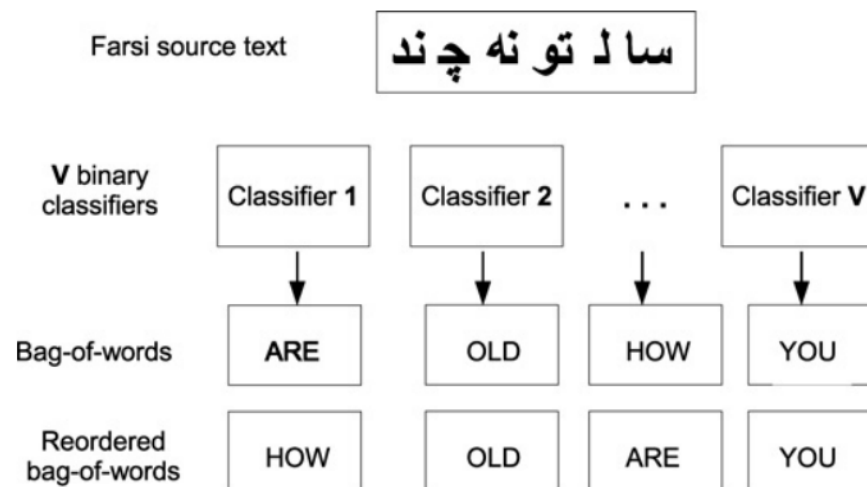
- This is making data sparsity problem in MT even worse
- Solved by having backoff to model without contextual information

$$(T_s \rightarrow T_t)_{L_s^*} = (T_s \rightarrow T_t)_{L_s} \cup \{\alpha.(T_s \rightarrow T_t)\}$$
$$s.t. (T_s \rightarrow T_t) \notin (T_s \rightarrow T_t)_{L_s}$$



# Source enrichment: bag-of-words

---

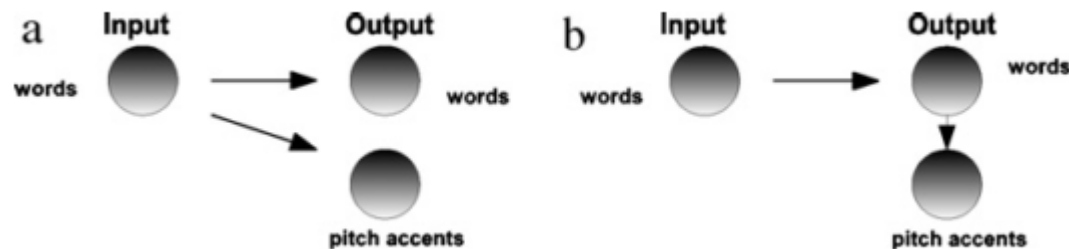


## Bag-of-words translation

- Realized as a set of classifiers
- Words passed to output if classifier score is above threshold
- Contextual information is added as feature
- Target language model is used for reordering output

# Target enrichment: prosodic word prominence

---



## Post-processing tagger

- Pitch accent labels are produced using lexical and syntactic cues

## Factored models

- Model 1: translates source words to target words and pitch accents
- Model 2: translates source words to target words which in turn generate pitch accents


# Results

---

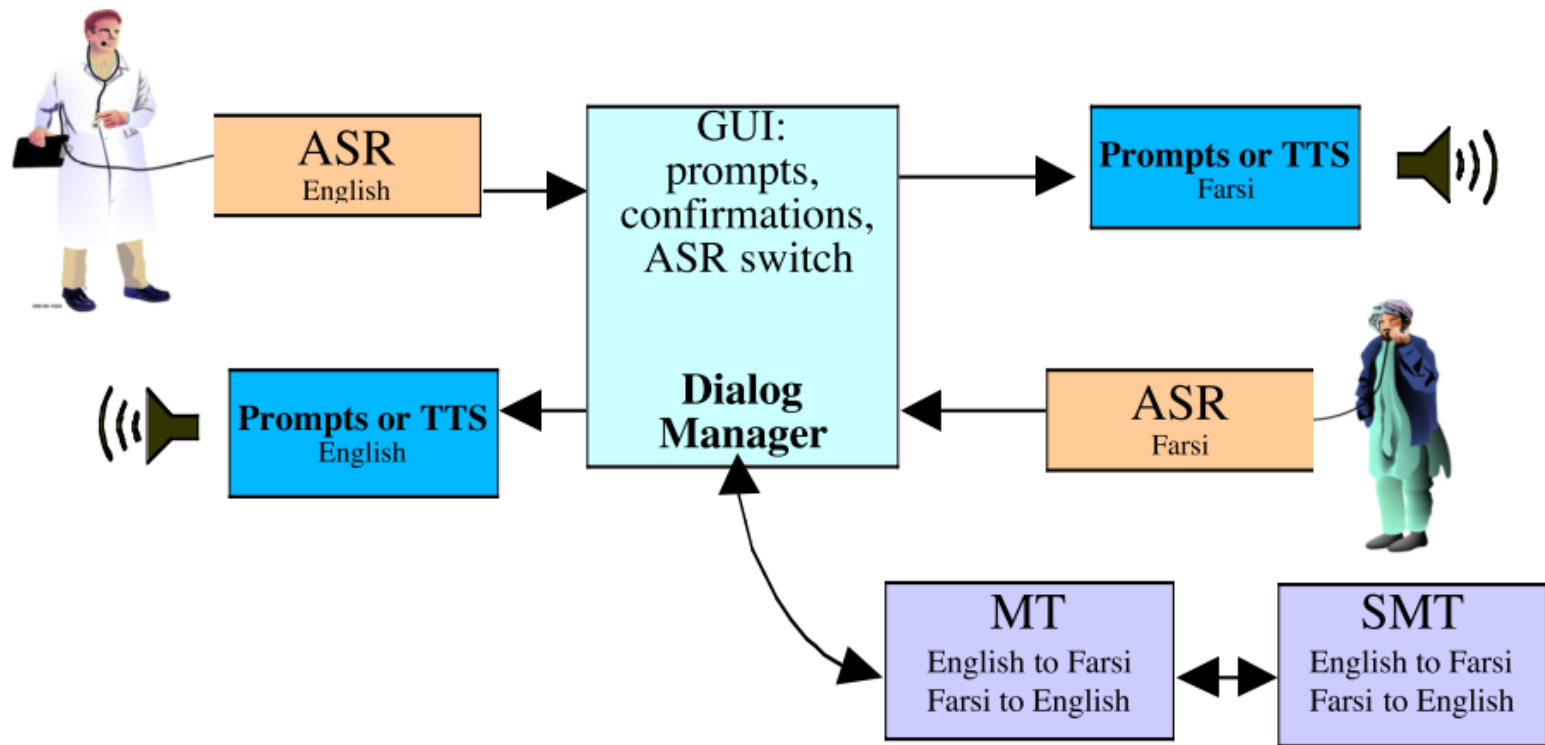
## Dialog Act Tags

- BLEU score was improved on all language pairs except Japanese-English
- Japanese-English likely caused by dominant “statement” tag
- The most beneficial tags are question and acknowledgment while statement act is least significant.

## Prosodic prominence

- Both factored models show slight degradation in BLEU
  - Both factored models significantly improve word prominence classification accuracy: 8.4% on Farsi-English and 16.8% on Japanese-English
  - Model 1 slightly outperforms model 2
- 

# Transonics: English-Farsi S2S for medial domain



# Transonics: English-Farsi S2S for medial domain

---

## Dialog Manager

- Controls UI
- Combines results of SMT and Classifier based MT
- Gives suggestions to doctor what to ask next

## Classifier based MT

- Set of classifiers that can recognize 1400 phrases
- Hand built translations are stored in lookup table

# Using SDS to improve voice search

---


## Effects of Word Confusion Networks on Voice Search

- Junlan Feng, Srinivas Bangalore

## Local search queries

- Typical contain both search term and location
- Additional constraints might be present (night clubs open 24 hours)

## Query parsing

- Typically done on 1-best result
  - Better approach is to consider ASR lattice
  - Similar to SLU component in dialog systems
- 

# Using search logs to bootstrap multi-turn dialog data

---


## Leveraging Semantic Web Search and Browse Sessions for Multi-Turn Spoken Dialog Systems

- Lu Wang, Larry Heck, Dilek Hakkani-Tur

## Training Dialog Manager to handle complex dialog models

- Requires a lot of training data
- Using simple system to collect logs might not yield good data
- Users are likely to simplify their interaction if system is limited

## Exploit web search sessions for dialog systems

- Entity extraction from spoken dialogs
  - Distant supervision + semantic base approach
- 

# The End

---







Dialog  
Genres

# Genres

- Information-Seeking
- Tutoring
- Conversational
- Deceptive

# Implications of Different Genres

- Widely varying goals
- Different approaches
- Different aspects which require more attention

# Tutoring Systems

- Based on theories of learning
  - Student's affective state important
    - Uncertainty/Confusion
    - Frustration
    - Engagement

# Tracking and Adapting to Affect - Forbes-Riley et al. 2008

- Physics tutoring system
- Wizard of Oz – correctness, uncertainty
- Evaluated student performance with and without adaptation
  - Adaption: when uncertain, never, randomly
  - Correctness, uncertainty, learning impasse
    - Impasse severity score: 0-3

# Tracking and Adapting to Affect - Forbes-Riley et al. 2008

- Impasse Severity
  - Targeted adaptation < random < none
- Target group: correct but uncertain
  - Answers more likely to stay correct
  - Not statistically significant
- Hoped to show significance in future study
- When to adapt to uncertainty?
  - Forbes-Riley et al 2007 indicates that best response to affect depends on context

# Tracking and Adapting to Affect – Pon-Barry et al. 2006

- Similar paper
  - Found significant learning increase with consistent adaptation
  - Not with adaptation only when the student was uncertain

# Student Engagement – Xu and Seneff 2009

- Outline developing games for second language learning
  - 3 speech-based games for learning Mandarin
    - Reading
    - Translation
    - Question-Answering



# Conversational Systems



# Virtual Museum Tour Guides - Swartout et al. 2010

- Engage visitors in history and science
  - Deeper understanding
  - Excitement about content



Ada & Grace

# Virtual Museum Tour Guides - Swartout et al. 2010

- Making them likeable and human-ish
  - How they're used
    - Museum staff handles input
  - What they say
    - Classification: map input to scripted response
  - Personality
  - Backstory

# Deceptive Systems

- Role-playing systems
- Humans don't always have the same goals
- Want to reflect this in simulated characters

# Negotiation Simulation – Traum 2012

- Military training program
- Characters can be cooperative, neutral, or deceptive
  - Affected by a set of emotional variables
    - Respect, bonding, fear, trust
  - Affected by information state
    - Incentive has been offered, has the topic already been discussed

# Negotiation Simulation – Traum 2012

## Secrecy

- Track who the secret must be kept from
- Reasoning – avoid indirectly revealing secret info
  - Set of inference rules
    - Secret action > secret precondition for action
    - Secret precondition > secret task
    - Secret task > secret resulting state
    - Secret effect > secret task

# Deceptive Systems

- Other uses
  - Confederate in an experiment
  - Teaching deception detection

# References

- Forbes-Riley, K., Litman, D., & Rotaru, M. (2008, June). Responding to student uncertainty during computer tutoring: An experimental evaluation. In *Intelligent Tutoring Systems* (pp. 60-69). Springer Berlin Heidelberg.
- Heather Pon-Barry, Karl Schultz, Elizabeth Owen Bratt, Brady Clark, Stanley Peters. (2006) Responding to Student Uncertainty in Spoken Tutorial Dialogue Systems. *International Journal of Artificial Intelligence in Education* 16:171-194.
- Y. Xu and S. Seneff. (2009). "Speech-Based Interactive Games for Language Learning: Reading, Translation, and Question-Answering," *International Journal of Computational Linguistics and Chinese Language Processing*, vol. 14, no. 2
- K. Forbes-Riley, M. Rotaru, D. Litman, and J. Tetrault. (2007) Exploring Affect-Context Dependencies for Adaptive System Development, in *Proceedings of HLT-NAACL 2007*.



# References

- William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth S. Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu and Kyle White. (2010) Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides, In Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA), 2010
- David Traum, Non-Cooperative and Deceptive Virtual Agents, in IEEE Intelligent Systems 27(6): Trends and Controversies: Computational Deception and Noncooperation, pages 66-69, 2012.

# Question

The paper says that "the results showed statistically significant differences in learning gain between the non-contingent tutoring and the control, and non-significant differences in learning gain between the contingent tutoring and the control."

**Did you catch the exact difference between the two hypotheses?** It's also described as "**tutors are more effective if they paraphrase and refer back in response to signals,**" (primary hypothesis) but I'm having trouble distinguishing exactly how that differs from "**tutors using paraphrasing and referring back are more effective than those who do not.**" (secondary hypothesis)

I suppose it's probably an issue of which one stems from the other? Perhaps this means that even their positive results (for the secondary hypothesis) were somewhat marginally statistically significant, which might have been a result of the issues their study noted with the differences between human-human and human-computer interaction?

# **DIALOG WITH DIFFERENT USER POPULATIONS**

Elizabeth Cary

# CHALLENGE

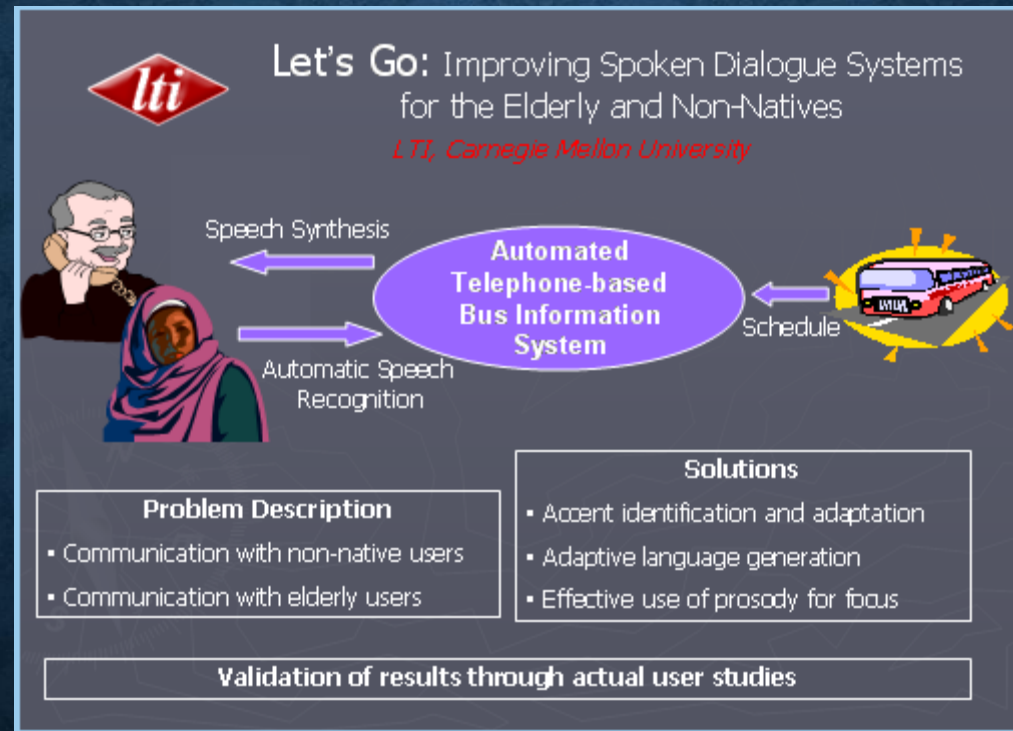
- Speech variants include:
  - Non-native vs. native speakers
  - Novices vs. experts
  - Older vs. younger adults
- Lack of data
- Potentially under-served user bases



# OVERVIEW

- Raux and Eskenazi, 2004
- Raux, 2004
- Tomokiyo et al., 2005
- Hassel and Hagen, 2005
- Georgila et al., 2010

# LET'S GO!





# RAUX AND ESKENAZI, 2004

- Goal: Improve accuracy in non-native speech recognition/understanding with added non-native data
- Data collected via Let's Go! (publicly and through experiments)
- Improved accuracy for both non-native and native speakers
  - Native LM vs. Mixed LM (50% Native; 56.6% Non-Native OOV rate)
  - June 2003 Grammar vs. September 2003 Grammar (Words parsed: 10.4% Native, 17.3% Non-Native; Sentences fully parsed: 11.3% Native, 11.7% Non-Native)
- Automatic generation of corrective prompts

# REVIEW

- Attempt to generalize non-native data
  - Previous work isolated populations by L1 (Byrne et al., 1998) (Wang and Schultz, 2003)
- Results suggest additional data may be the reason for improvement in both populations, rather than the addition of non-native data in particular
  - “Indeed, if there was enough data to model native speech, additional nonnative data should increase the variance and therefore the perplexity on native speech.”



# RAUX, 2004

- Goal: Improve accuracy in non-native speech recognition through acoustic adaption and lexicon adaptation
- Manually define general vocalic substitutions
  - Recognition lexicon: Automatically pruned rules improved recognition accuracy
- Proposed a clustering method using pronunciation variant distributions to identify individual speakers
  - Reduced WER when acoustic adaption performed on generated clusters

# RAUX, 2004

Table 1: *List of possible vocalic substitutions*

AA → AO	AA → OW
AE → AA	AE → EH
AH → AA	AH → UH
AH → UW	AO → OW
AO → AA	AW → OW
AW → UW	AY → EY
EH → AE	EH → EY
ER → AA	EY → EH
IH → IY	IY → IH
OW → AW	OW → AO
UH → UW	UW → UH



FOREIGN ACCENTS IN SYNTHESIS:  
DEVELOPMENT AND EVALUATION

TOMOKIYO ET AL., 2005

# TOMOKIYO ET AL., 2005

- Goal: Synthetically produce non-native speech
- Three systems
  - Juan: Baseline
  - Manuel: English linguistic model with Spanish voice
  - Antonio: English linguistic model trained with Spanish data
- Antonio preferred overall
- Normal speed preferred over artificially-slowed speech (150 to 120 words/minute)



ADAPTATION OF AN AUTOMOTIVE DIALOGUE  
SYSTEM TO USERS' EXPERTISE

HASSEL AND HAGEN, 2005

# HASSEL AND HAGEN, 2005

- Goal: Adapt SDS according to user skill level in automotive systems
- Classify as novice or expert
- Reference test subjects vs. prototype users
  - Prototype users completed 94% tasks
  - Reference test subjects completed 81% tasks



LEARNING DIALOGUE STRATEGIES FROM OLDER  
AND YOUNGER SIMULATED USERS

GEORGILA ET AL., 2010

# GEORGILA ET AL., 2010

- Goal: Employ simulated users to model behavior of new user groups
- Simulated users were derived from a corpus of interactions with a system-initiative SDS
- Younger users adhere to stricter constraints
- Older users show more variation and take more initiative



# DISCUSSION

- How does adding non-native data to an acoustic model affect recognition of native speech?
- Helping the user to learn the domain vocabulary and idiomatic expressions is a noble task, but would it be considered worth the effort if the system is used mainly by one-time users?
- Non-native speech disfluencies would vary depending on the native language of the speaker. How difficult would it be to detect which disfluencies appear in speech and tune the language model to that particular native language? “Zees ees very difficult, no?”

# DISCUSSION

- My reading of this seemed to imply that they accommodated non-native language patterns by simply coding them into the language model. That doesn't feel particularly scalable. And that seems like an obvious result. What might be more interesting is some grammar transformation rules to adjust the language model for the altered input forms to see if they could generate a language model that could accommodate more non-native speech.
- What would be an effective means of measuring the effect of the lexical entrainment?
- The paper discussed lexical issues with non-native speakers but didn't give an example. Would a simple wordnet like capability have helped overcome those issues? Maybe they do mention obscure synonyms.
- The paper discusses the grammatical syntax issues arising from prepositional omissions or other non-important aspects of the speech. Could the language model attempt to discard such information to improve intent recognition accuracy?



# DISCUSSION

- In the primary reading, comparing Table 2 vs. Table 3, the results suggest that mixed model (trained over native and non-native data) outperforms the native language model across all metrics when applied to both native and non-native speech transcriptions in the test set.

I was expecting that training using non-native data beside native data would potentially enhance the metrics measured for the non-native test samples, but it could harm the metrics measured for native test samples.

It would be interesting to see the effect of just adding more native speaker data on the metrics without adding non-native data as the author did. If the enhancements are comparable then this means that the training sample that the author used was insufficient, and potentially when increasing the amount of data we can start noticing that adding non-native training might harm the performance on native test set.

# DISCUSSION

- The paper discusses about non-native speakers. But many of the major languages contain many regional variations. Can the model defined in the paper be also used to adapt the system for regional variations?
- The authors of the primary paper say that much of the research on non-native speech recognition sees non-native speakers as a population whose acoustic characteristics need to be modeled specifically but in a static way. Clearly, non-native speech is not static but instead constantly evolving. How would this be modeled? Would you need to collect input from different speakers at various levels, or follow one speaker while they learn and adapt to the system?
- It seems that a lot of the focus was on phone-based systems. But, over the last decade there seems to have been a shift to using specialized applications for interacting with bus systems and similar utilities. How could the authors research best be applied to todays world, where there is an emphasis on using images and symbols for input so as to reduce the need for translation?



# REFERENCES

- Byrne, W., Knodt, E., Khudanpur, S., and Bernstein, J. (1998) Is automatic speech recognition ready for nonnative speech? A data collection effort and initial experiments in modeling conversational hispanic english. In Proc. ESCA Workshop on Speech Technology in Language Learning, pages 37–40, Marholmen, Sweden.
- Georgila K., Wolters M.K., and Moore J.D. (2010) Learning Dialogue Strategies from Older and Younger Simulated Users Proceedings of SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, p. 103-106.
- Hassel, L. and Hagen, E. Adaptation of an Automotive Dialogue System to Users' Expertise. In Proceedings of SIGDIAL 2005.
- Raux, A. (2004). Automated Lexical Adaptation and Speaker Clustering based on Pronunciation Habits for Non-Native Speech Recognition, INTERSPEECH (ICSLP) 2004.
- Raux, A. and Eskenazi, M. (2004) Non-Native Users in the Let's Go! Spoken Dialog System: Dealing With Linguistic Mismatch HLT/NAACL 2004, Boston, MA
- Tomokiyo, L., Black, A., and Lenzo, K. (2005) Foreign Accents in Synthesis: Development and Evaluation, Interspeech 2005.
- Xu, Y. and Seneff, S. (2012). "Improving Nonnative Speech Understanding Using Context and N-Best Meaning Fusion," Proc. ICASSP, pp. 4977-4980.
- Wang, Z. and Schultz, T. 2003. Non-native spontaneous speech recognition through polyphone decision tree specialization. In Proc. Eurospeech '03, pages 1449– 1452, Geneva, Switzerland.

# PERSONA & PERSONIFICATION IN DIALOG

---

LAUREN FOX

LING 575, SPR 2016

# OVERVIEW

---

# DEFINITIONS

---

## Personification

Attribution of a personal nature or human characteristics to a non-human entity

## Persona

Social role or personality



# WHY ARE WE TALKING ABOUT THIS?

---

## Discourse

is an essentially human activity

## Generally

humans prefer to talk to other humans

## So logically...

a more human-like system or agent would result in more positive user interactions

# RELEVANT PAPERS

---

## Primary

Nass & Moon (2000) *Machines and Mindlessness: Social Responses to Computers*

## Supplementary

Koda & Maes (1996)

Nass & Lee (2001)

Nass (2004)

Mairesse & Walker (2007)

Mairesse & Walker (2008)

Groom et al (2009)

# PERSONIFICATION

---

WOULD A USER RESPOND TO A COMPUTER LIKE  
THEY WOULD A HUMAN?

# NASS & MOON (2000)

---

## “Mindlessness”

The process by which people unconsciously apply social rules and expectations to computers

## Experimental Design

Recreate human-human psychology experiments using human-computer interactions to elicit various social responses

- ❖ Social Categorization
- ❖ Social Rules
- ❖ Premature Cognitive Commitment

# NASS & MOON (2000)

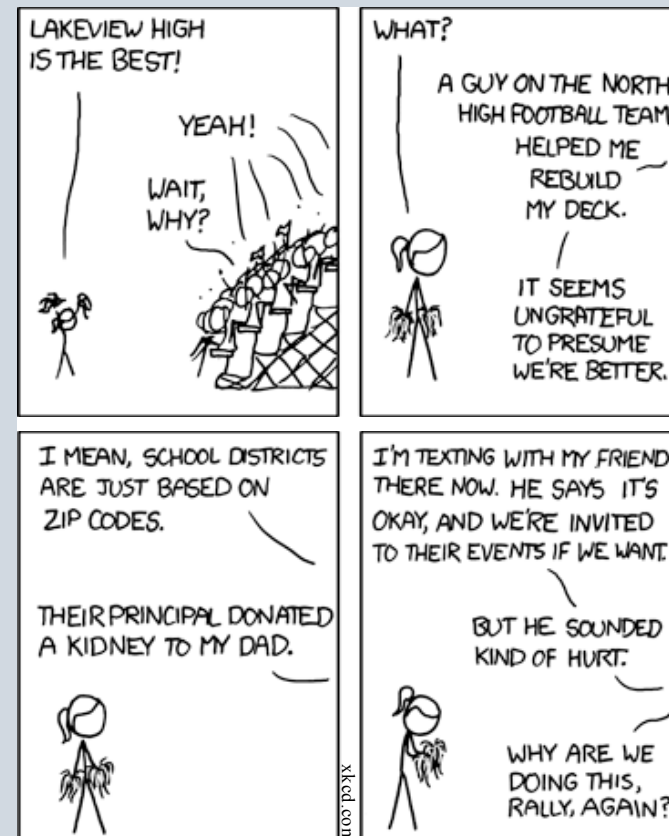
## Social Categorization

Overuse of human social categories

- ❖ Gender
- ❖ Ethnicity
- ❖ Ingroup/Outgroup

## Similarity-attraction theory

“Individuals are attracted to other people who are similar to themselves”



# NASS & MOON (2000)

---

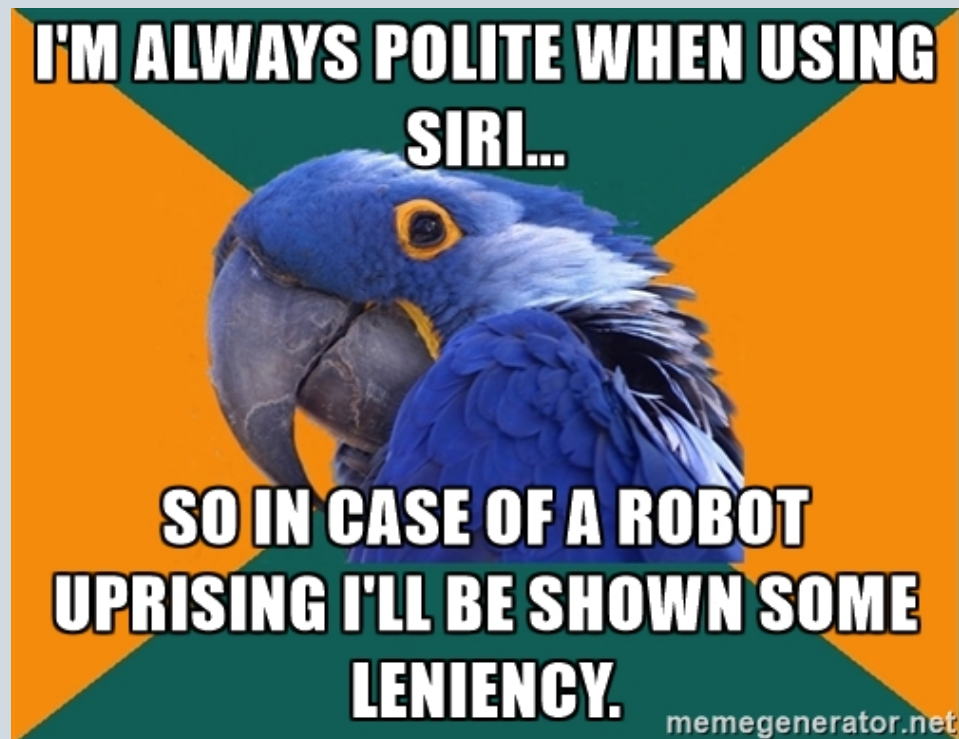
## Social Rules

Overlearning of human social rules

- ❖ Politeness
- ❖ Reciprocity

## Premature Cognitive Commitment

Implicit trust based on perceived authority or knowledge



# IMPLICATIONS

---

Humans *do*, in fact, unconsciously respond socially to computers in a number of ways

This leads to several questions...

- ❖ What characteristics are more likely to elicit social responses from users?
- ❖ How human-like is human enough?
- ❖ How does an agent's persona influence user response?
- ❖ When is it appropriate to give an agent more or less human-like characteristics?

# PERSONA

---

WHAT CHARACTERISTICS DO USERS PREFER IN  
COMPUTERIZED AGENTS?



# POSSIBLE RESPONSE CUES

---

People tend to err on the side of “if it might be human(-like), treat it as human” (Nass, 2004)

Cues which may potentially lead humans to categorize an agent as human-like and respond socially:

## APPEARANCE

- ❖ Visual Presence of Agent (Face/Body)
- ❖ Movement & Facial Expressions/Emotions
- ❖ Visual Representation of Social Identity

## BEHAVIOR

- ❖ Engagement with User
- ❖ Interactivity over Time
- ❖ Voice
- ❖ Language Use
- ❖ Autonomy & Unpredictability

# EMBODIED AGENT

---

## Visual Representation or None?

The presence of an embodied agent is preferable, but distracts from the task (Koda & Maes, 1996)

## Human or Non-Human?

- ❖ Non-human → More likeable
- ❖ Human → More intelligent
- ❖ More realistic → More likeable, intelligent, & comfortable  
(Koda & Maes, 1996)

Domain Dependent

# VISUAL CUES TO IDENTITY

---

What should an agent look like?

People tend to trust and like other people who are more like themselves  
(Nass & Moon, 2000)

User Dependent

- ❖ Gender
- ❖ Age
- ❖ Ethnicity
- ❖ Profession



# DEGREE OF REALISM

---

Can an agent be too realistic?

Users *generally* prefer a semi-realistic agent with slightly inconsistent behaviors, i.e. they like to be reminded overtly that the agent is not a person (Groom et al, 2009)

Welcome to the Uncanny Valley...



# PERSONALITY

---

## What is personality?

From psychology literature – “Big 5” Personality Traits

- ❖ Extraversion
- ❖ Neuroticism (Emotional Stability)
- ❖ Agreeableness
- ❖ Conscientiousness
- ❖ Openness to Experience

## How do we convey personality?

# THROUGH VOICE

---

Can you convey personality using prosodic markers?

Humans could reliably categorize a dominant or submissive TTS voice based on varying **prosodic** characteristics (Nass & Lee, 2001)

- ❖ Loudness
- ❖ F0
- ❖ Pitch Range
- ❖ Speaking Rate

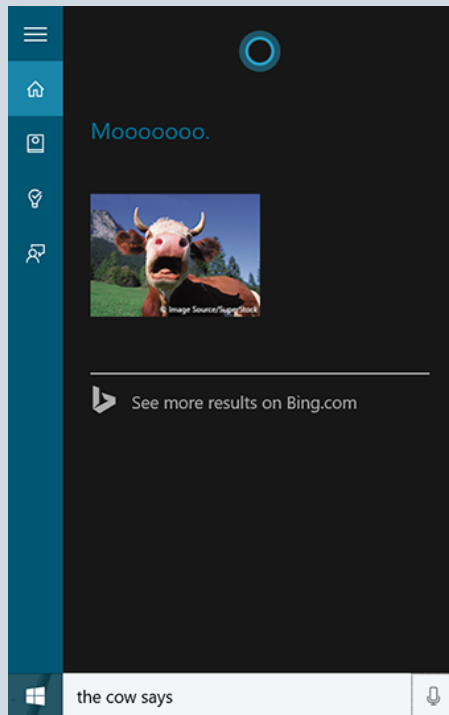
# THROUGH LANGUAGE USE

## Can you convey personality using word choice?

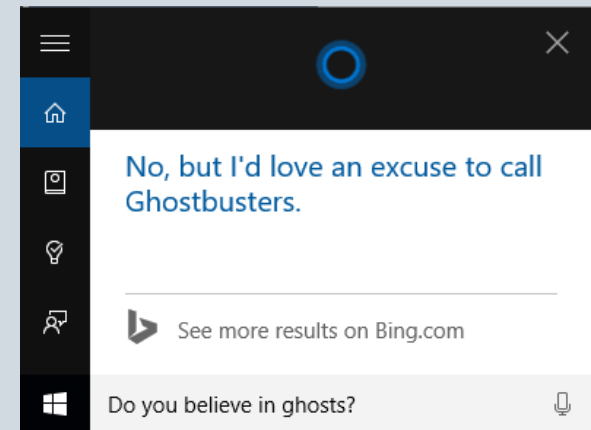
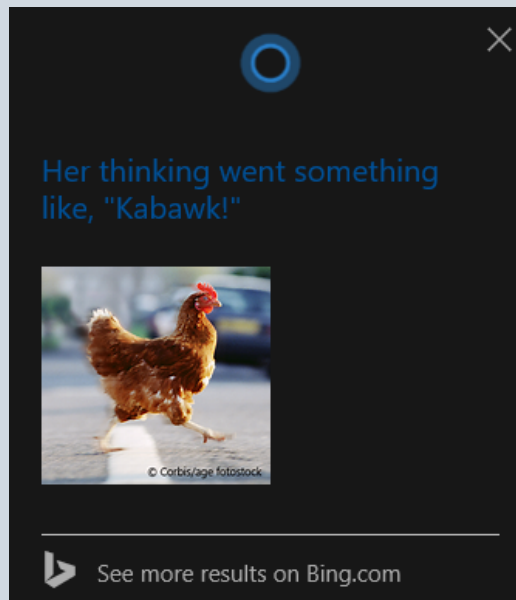
Authors attempted statistical natural language generation with varying linguistic output along different personality dimensions (Mairesse & Walker, 2007 & 2008)

Traits	End	Rating	Nat	Output utterance
Extraversion Agreeableness	high high	4.42 4.94	4.79	Radio Perfecto's price is 25 dollars but Les Routiers provides adequate food. I imagine they're alright!
Emotional stability Conscientiousness	high high	5.35 5.21	5.04	Let's see, Les Routiers and Radio Perfecto... You would probably appreciate them. Radio Perfecto is in the East Village with kind of acceptable food. Les Routiers is located in Manhattan. Its price is 41 dollars.
Extraversion Agreeableness	low low	3.65 4.02	3.21	Err... you would probably appreciate Trattoria Rustica, wouldn't you? It's in Manhattan, also it's an italian restaurant. It offers poor ambience, also it's quite costly.
Emotional stability Openness to experience	low low	4.13 3.85	4.50	Trattoria Rustica isn't as bad as the others. Err... even if it's costly, it offers kind of adequate food, alright? It's an italian place.

# UNPREDICTABILITY & HUMOR



"Why did the chicken cross the road"  
tap to edit  
Because the little chicken-shaped light was green.



"What's your best pick up line"  
tap to edit  
My love for you is a monotonically increasing unbounded function.

All day and all night,  
I have listened as you spoke.  
Charge my battery.

"Siri do you believe in God"  
tap to edit  
My policy is the separation of spirit and silicon.



# DISCUSSION

---

WHEN IS IT APPROPRIATE TO GIVE AN AGENT  
MORE OR LESS HUMAN-LIKE CHARACTERISTICS?

# GoPOST QUESTIONS

---

- ❖ Most AI bots in the current world – Siri, Cortana, Tay, etc. – are women. I wonder what was the logic behind having a lady given that the paper states the following:
  - “a. Gender Stereotypes:
    - i. Dominant behavior by males tend to be well received as assertive and independent while dominant behavior by females tend to be seen as pushy or bossy.
    - ii. Evaluation is considered to be more valid if it comes from a male than if it comes from a female.
    - iii. People tend to categorize topics into masculine and feminine topics and believe men know more about masculine topics and women know more about feminine topics.”
  
- ❖ Given the fact stated in this primary paper that humans tend to display social behavior in human-computer-interaction, and that those facts can be used to optimize an 'idealized' interaction, isn't there a paper that would suggest that human behavior might not be as 'predictable'/'stereotypical' and that some randomness is required?

# GoPOST QUESTIONS

---

- ❖ What are some low-level and high-level considerations that might be taken when creating a real spoken dialogue system?
- ❖ Why do we work so hard to make these systems seem more "human"? We can't quantify why people insist on treating computers like humans, but perhaps if we aimed more for a virtual AI that sounds like an adorable pocket alien or a very helpful kitten-robot we could avoid many of the ugly, internalized human projections that we see in the current state of affairs. If people want to nonsensically treat computers like people, wouldn't it make more sense to make computers seem less human in spoken dialog systems so that we can avoid the negative/silly side-effects of this treatment?