

# CS 124/LINGUIST 180

## From Languages to Information

Dan Jurafsky  
Stanford University

# Conversational Agents

# Conversational Agents

## AKA Dialog Agents

Phone-based Personal Assistants

SIRI, Alexa, Cortana, Google Assistant

Talking to your car

Communicating with robots

Clinical uses for mental health

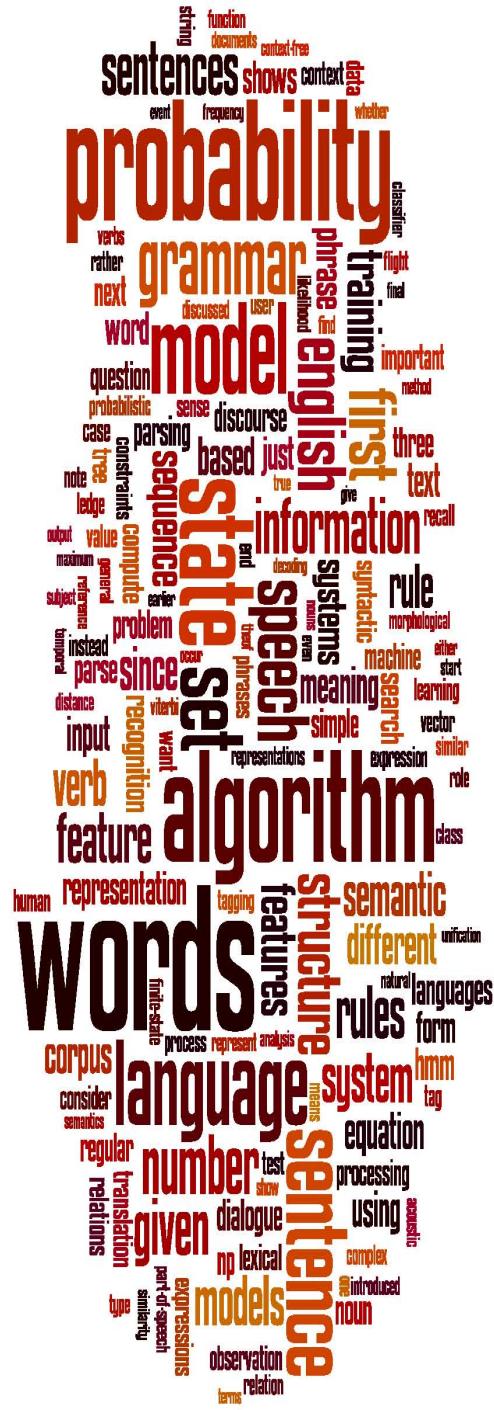
Chatting for fun

# Two kind of conversational agents

1. Chatbots
2. (Goal-based) Dialog agents

- *SIRI, interfaces to cars, robots,*
- *booking flights or restaurants*

Recently I've noticed that the word "chatbots" is sometimes used in the popular press for both.  
I'll use it only for #1.



# CS 124/LINGUIST 180

## From Languages to Information

Dan Jurafsky  
Stanford University

# Properties of Human Conversation

# Linguistic Background: Properties of Human Conversation

- **Turn-taking**
  - When to take the floor?
  - When to yield the floor?
- **End-pointing**
  - In speech, deciding whether the user has stopped talking.

# A human-human conversation

- C<sub>1</sub>: ...I need to travel in May.
- A<sub>1</sub>: And, what day in May did you want to travel?
- C<sub>2</sub>: OK uh I need to be there for a meeting that's from the 12th to the 15th.
- A<sub>2</sub>: And you're flying into what city?
- C<sub>3</sub>: Seattle.
- A<sub>3</sub>: And what time would you like to leave Pittsburgh?
- C<sub>4</sub>: Uh hmm I don't think there's many options for non-stop.
- A<sub>4</sub>: Right. There's three non-stops today.
- C<sub>5</sub>: What are they?
- A<sub>5</sub>: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
- C<sub>6</sub>: OK I'll take the 5ish flight on the night before on the 11th.
- A<sub>6</sub>: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
- C<sub>7</sub>: OK.
- A<sub>7</sub>: And you said returning on May 15th?
- C<sub>8</sub>: Uh, yeah, at the end of the day.
- A<sub>8</sub>: OK. There's #two non-stops ... #
- C<sub>9</sub>: #Act... actually #, what day of the week is the 15th?
- A<sub>9</sub>: It's a Friday.
- C<sub>10</sub>: Uh hmm. I would consider staying there an extra day til Sunday.
- A<sub>10</sub>: OK...OK. On Sunday I have ...

# Language as Action



**Each turn in a dialogue is a kind of action**  
Wittgenstein (1953) and Austin (1962)

# Speech Acts (aka Dialog Acts)

**Constatives:** committing the speaker to something's being the case (*answering, claiming, confirming, denying, disagreeing, stating*)

**Directives:** attempts by the speaker to get the addressee to do something (*advising, asking, forbidding, inviting, ordering, requesting*)

**Commissives:** committing the speaker to some future course of action (*promising, planning, vowing, betting, opposing*)

**Acknowledgments:** express the speaker's attitude regarding the hearer with respect to some social action (*apologizing, greeting, thanking, accepting an acknowledgment*)

# Speech acts

"Turn up the music!"

Directive

"What day in May do you want to travel?"

Directive

"I need to travel in May"

Constative

Thanks

Acknowledgement

Grounding refers to establishing the time, location or actuality of a situation according to some reference point

# Grounding

- Why do elevator buttons light up?
- Clark (1996) (after Norman 1988)

**Principle of closure.** Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it

- What is the linguistic correlate of this?

# Grounding

A: And you said returning on May 15th?

C: Uh, yeah, at the end of the day.

A: **OK**

C: OK I'll take the 5ish flight on the night before on the 11th.

A: **On the 11th? OK.**

C: ...I need to travel in May.

A: **And**, what day **in May** did you want to travel?

# Grounding and Confirmation

- We need to know whether an action succeeded or failed
- Talking is an action!
- I need to know if my action succeeded
  - i.e. the hearer understood my turn!

# What happens if the system doesn't ground?

Cohen et al (2004)

System: Did you want to review some more of  
your personal profile?

Caller: No.

**Bad!**

System: What's next?

System: Did you want to review some more of  
your personal profile?

Caller: No.

**Good!**

System: Okay, what's next?

# Subdialogues

Agent... OK. There's #two non-stops#

Client: #Act. . . actually#, what day of the week is the 15th?

Agent: It's a Friday.

Client: Uh hmm. I would consider staying there an extra day til Sunday.

Agent: OK...OK. On Sunday I have ...

# Clarification Subdialogs

User: What do you have going to UNKNOWN WORD on the 5th?

System: Let's see, going where on the 5th?

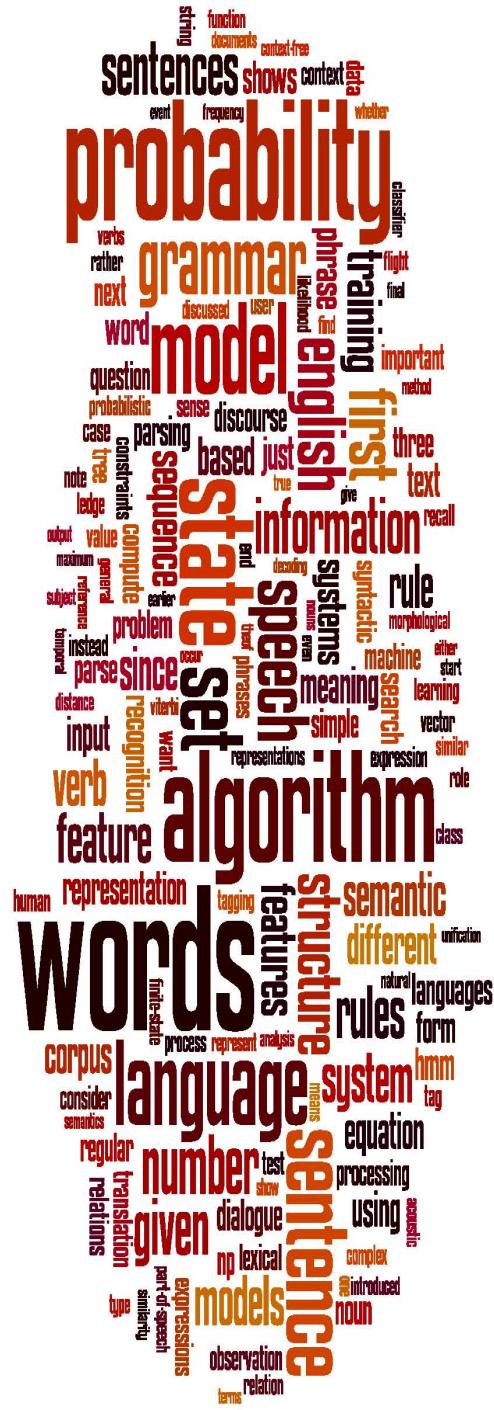
User: Going to Hong Kong.

System: OK, here are some flights...

## Even harder problems: Inference

- Agent: And, what day in May did you want to travel?

Client: OK uh I need to be there for a meeting that's from the 12th to the 15th.



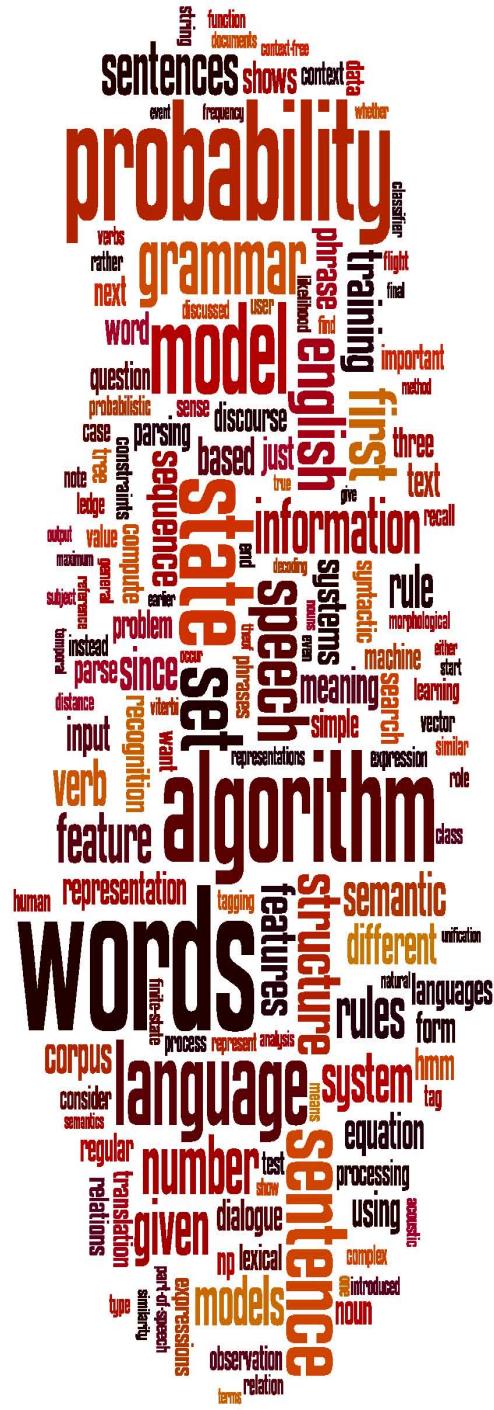
# CS 124/LINGUIST 180

## From Languages to Information

# Dan Jurafsky

# Stanford University

# Properties of Human Conversation

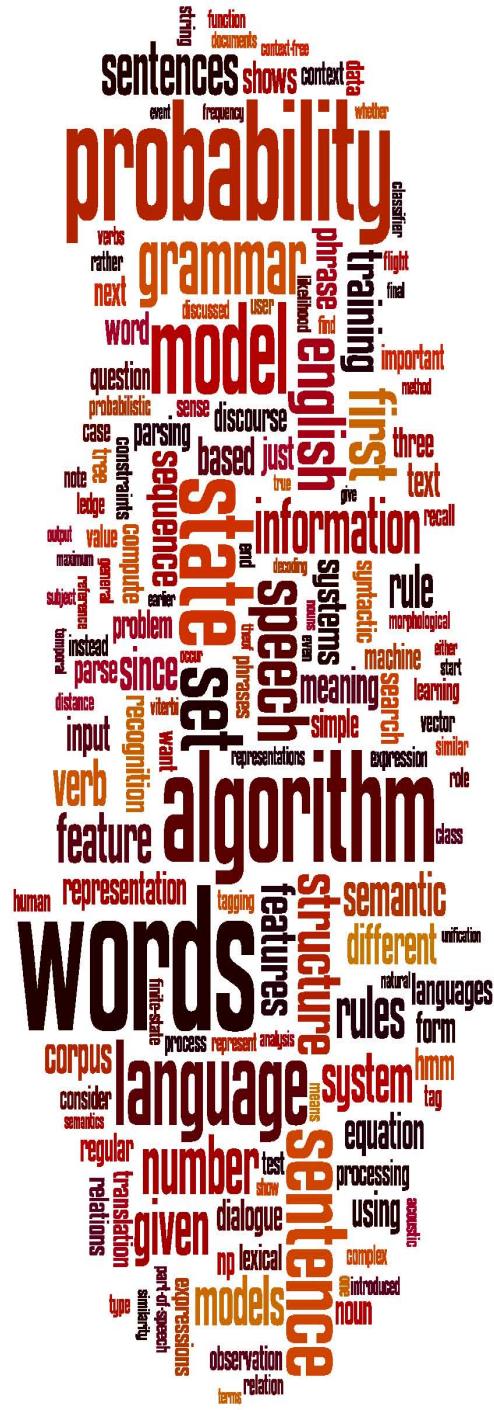


# PA 6

# Foreshadowing

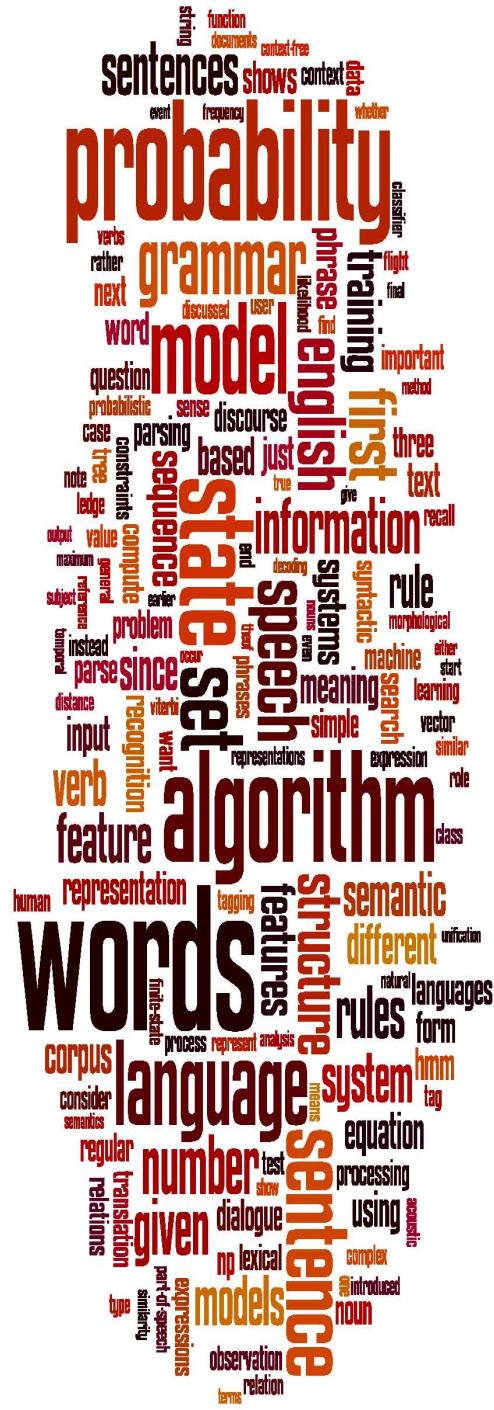
# PA 6: Chatbot!

- Every group will build a movie recommending chatbot!
  - Hence: Chatbot + Recommendation Engine
- We'll give you the movie ratings database of 9125 movies rated by 671 users
- There will be starter mode + creative mode!
- Bureaucratic details:
  - Late days CANNOT be used on this assignment.
  - You must work in groups of 2-4 members
  - All submissions will be graded according to the same criteria, regardless of group size.



# PA 6

# Foreshadowing



# Rule-based Chatbots: ELIZA

# Chatbots!

- ELIZA (1966)
- PARRY (1968)
  - The first system to pass the Turing test!!!!
- ALICE
- CLEVER
- Microsoft Little Bing 小冰

# Chatbot Architectures

## **Rule-based**

1. Pattern-action rules (Eliza)  
+ a mental model (Parry)

## **Corpus-based (from large chat corpus)**

2. Information Retrieval
3. Neural network encoder-decoder

Chatbots

Sometimes for fun

But originally designed as a testbed for  
clinical therapy

(And still used that way!)

# Eliza: Weizenbaum (1966)

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

...

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

You are like my father in some ways.

**WHAT RESEMBLANCE DO YOU SEE**

You are not very aggressive but I think you don't want me to notice that.

**WHAT MAKES YOU THINK I AM NOT AGGRESSIVE**

You don't argue with me.

**WHY DO YOU THINK I DON'T ARGUE WITH YOU**

You are afraid of me.

**DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU**

My father is afraid of everybody.

**WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER**

Bullies.

**DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR**

How does it work?

Eliza uses this one weird trick

Person-centered therapy was developed by Carl Rogers in the 1940s. This type of therapy diverged from the traditional model of the therapist as expert and moved instead toward a nondirective, empathic approach that empowers and motivates the client in the therapeutic process. The therapy is based on Rogers's belief that every human being strives for and has the capacity to fulfill his or her own potential. Person-centered therapy, also known as Rogerian therapy, has had a tremendous impact on the field of psychotherapy and many other disciplines.

# That trick: be a Rogerian psychologist

Draw the patient out by reflecting patient's statements back at them

Rare type of conversation in which one can “assume the pose of knowing almost nothing of the real world”

# Rogerian psychologist

Patient: "I went for a long boat ride"

Psychiatrist: "Tell me about boats"

- You don't assume she didn't know what a boat is
- You assume she had some conversational goal
- Chatbots trying to pass the Turing test usually choose some such domain

# Eliza pattern/transform rules

( 0 YOU 0 ME) [pattern]

→

(WHAT MAKES YOU THINK I 3 YOU) [transform]

0 means Kleene \*

The 3 is the constituent # in pattern

You hate me

WHAT MAKES YOU THINK I HATE YOU

# Eliza Rules

keyword	pattern	Ranked list of transforms
$(K$	$((D_1)$	$(R_{1, 1})$
		$(R_{1, 2}) \dots (R_{1, m_1})$
	$((D_2)$	$(R_{2, 1})$
		$(R_{2, 2}) \dots (R_{2, m_2})$
	$\vdots$	$\vdots$
	$((D_n)$	$(R_{n, 1})$
		$(R_{n, 2}) \dots (R_{n, m_n}))$

# ELIZA: given *sentence* return *response*

- Find the word  $w$  in *sentence* that has the highest keyword rank
- If  $w$  exists:
  - Check each rule for  $w$  in ranked order
    - Choose first one that matches sentence
    - *response*  $\leftarrow$  apply transform
- Else
  - *response*  $\leftarrow$  apply "NONE" transform, or
  - *response*  $\leftarrow$  grab an action off the memory queue

Keywords are ranked from specific to general

I know everybody laughed at me

- “I” is a very general keyword:

I: (I \*) -> (You say you 2)

YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

- “Everybody” is much more interesting (someone using universals like everybody/always is probably “referring to some quite specific event or person”)

WHO IN PARTICULAR ARE YOU THINKING OF?

- Implementation: keywords stored with their rank

**Everybody** 5 (*transformation rules*)

**I** 0 (*transformation rules*)

NONE

PLEASE GO ON

THAT'S VERY INTERESTING

I SEE

# Memory

```
(MEMORY MY  
  (0 YOUR 0 = LETS DISCUSS FURTHER WHY YOUR  
3)  
  (0 YOUR 0 = EARLIER YOU SAID YOUR 3)
```

- Whenever “MY” is highest keyword
  - Randomly select a transform on the MEMORY list
  - Apply to sentence
  - Store on a stack
- Later, if no keyword matches a sentence

# Other Eliza stuff

- Rules can refer to classes of words
  - Family = mother, father, brother, sister
  - NOUN = ...
- Don't reuse transforms in the same conversation
  - Whenever we use a transform associated with a pattern
  - We increment a counter for that rule
  - So the next time we use the next ranked transform
- Some basic transforms happen during input processing
  - I -> YOU
  - YOU -> I

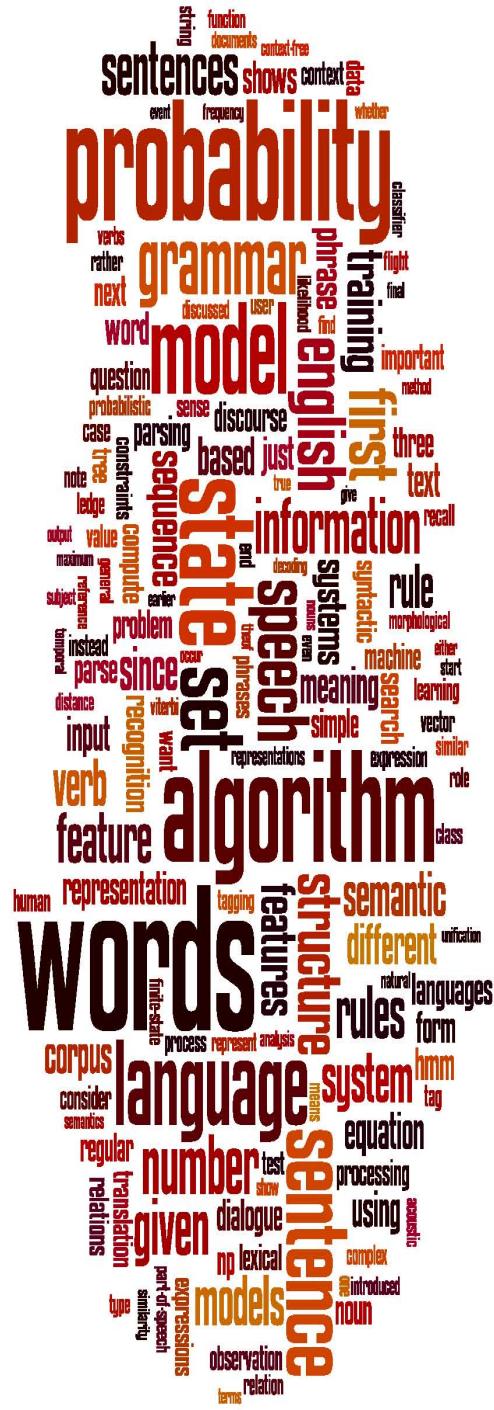
# Ethical implications: Anthropomorphism and Privacy

- People became deeply emotionally involved with the program
- Weizenbaum's secretary asked him to leave the room when she talked with ELIZA
- When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications
  - Suggesting that they were having quite private conversations with ELIZA

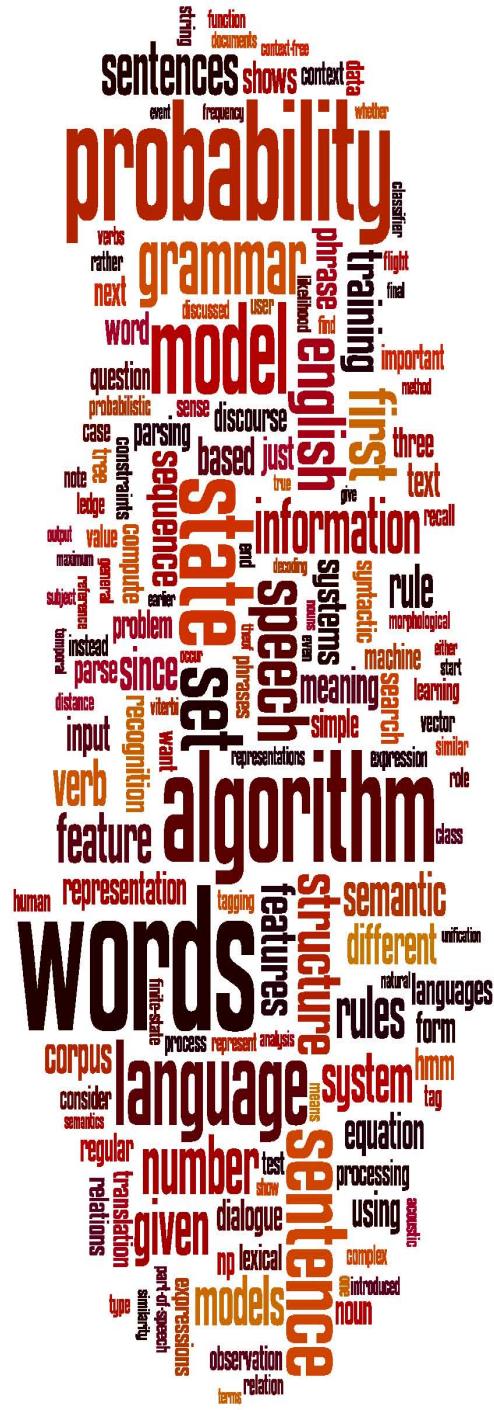
# Ethical implications: Anthropomorphism

Anthropomorphism and the Heider-Simmel Illusion

<https://www.youtube.com/watch?v=8FIEZXMUM2I>



# Rule-based Chatbots: ELIZA



# Rule-based Chatbots: PARRY

# Parry

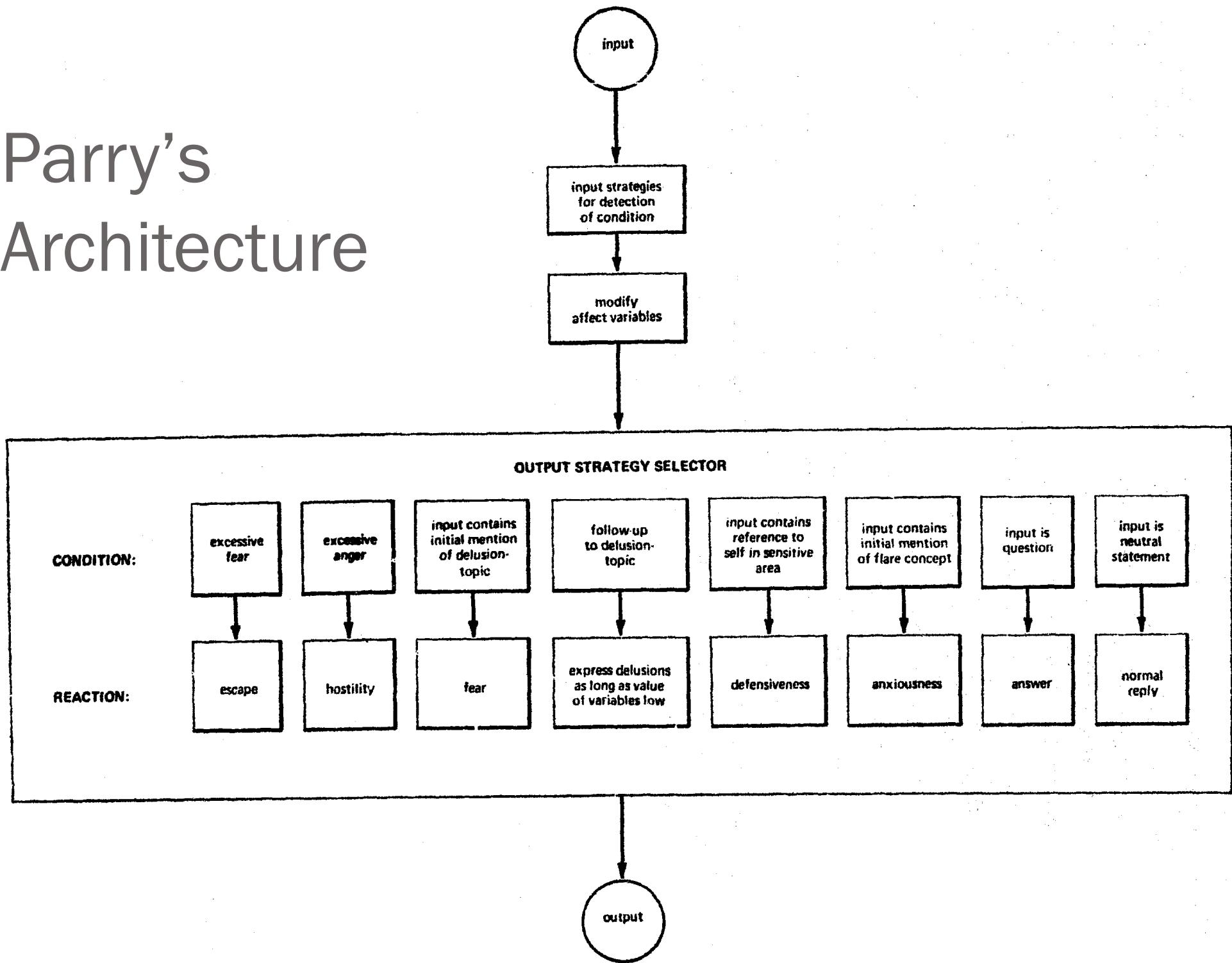


- Colby 1971 at Stanford
- Same pattern-response structure as Eliza
- But a much richer:
  - control structure
  - language understanding capabilities
  - mental model: Parry has affective variables
    - Anger, Fear, Mistrust
    - “If Anger level is high, respond with hostility”
- The first system to pass the Turing test (in 1971)
  - Psychiatrists couldn't distinguish interviews with PARRY from (text transcripts of) interviews with real paranoids

# Parry's persona

- 28-year-old single man, post office clerk
- no siblings and lives alone
- sensitive about his physical appearance, his family, his religion, his education and the topic of sex.
- hobbies are movies and gambling on horseracing,
- recently attacked a bookie, claiming the bookie did not pay off in a bet.
- afterwards worried about possible underworld retaliation
- eager to tell his story to non-threatening listeners.

# Parry's Architecture



# Affect variables

- Fear and Anger (each ranging 0-20)
- Mistrust (ranging 0-15)
- Initial conditions: All low
- After each user turn, if nothing malevolent in input
  - Anger drops by 1, Fear drops by 0.3
  - Mistrust drops by 0.05 to base level
- Otherwise depends on what the user says
  - Each user statement can change Fear and Anger
    - Insults increases Anger by some percentage
  - Mistrust goes up if Fear or Anger do

# Lots of complex I-O rules

- User implies Parry is mentally ill
  - Rise in Fear and Anger
- User mentions “Mafia” or associated concepts (“kill”):
  - First mention: rise in Fear
  - Later mentions: depends on willingness to discuss, which depends on current levels of Fear, Anger, Mistrust
- User mentions Parry
  - Flattery (positive mention)
    - Decreases fear/anger if Mistrust is low
    - Increases Anger if Mustrust is high
  - User attitudes toward Parry
    - Negative attitudes (fear, disbelief) increase Fear/Anger

# Flare concepts

- List of concepts related to Mafia
- An ordered graph designed to lead interviewer to topic  
horses → horseracing → gambling → bookies → underworld → Mafia
- The mention of a new flare topic by interviewer causes a rise in Fear
- Flare topics cause Parry to give preset responses to that flare

# Each sentence is mapped into a conceptualization

- A predication on a conceptual object
- A predication on a relation between two objects
- A predication on an attribute:

What is your work?

What sort of work do you do?

Where do you work? → (your work?)

What do you do for a living?

What is your job?

Do you have a job?

What is your occupation

- Complex Pattern/transform rules

- Different predicates (fear, afraid of)
- Ordering (You are afraid of me = I frighten you)

# Detecting Other's Intent

恶意的 行善的  
( OTHER'S INTENTION ) ← < MALEVOLENCE > | < BENEVOLENCE > | < NEUTRAL >

## MALEVOLENCE-DETECTION RULES

1. < malevolence > ← < mental harm > | < physical threat >
2. < mental harm > ← < humiliation > | < subjugation >
3. < physical threat > ← < direct attack > | < induced attack >
4. < humiliation > ← < explicit insult > | < implicit insult >
5. < subjugation > ← < constraint > | < coercive treatment >
6. < direct attack > ← CONCEPTUALIZATIONS ([you get electric shock], [are you afraid mafia kill you?])
7. < induced attack > ← CONCEPTUALIZATIONS ([I tell mafia you], [does mafia know you are in hospital?])
8. < explicit insult > ← CONCEPTUALIZATIONS ([you are hostile], [you are mentally ill?])
9. < implicit insult > ← CONCEPTUALIZATIONS ([tell me your sexlife], [are you sure?])
10. < constraint > ← CONCEPTUALIZATIONS ([you stay in hospital], [you belong on locked ward])
11. < coercive treatment > ← CONCEPTUALIZATIONS ([I hypnotize you], [you need tranquilizers])

# Detecting Other's Intent

---

$\langle \text{OTHER'S INTENTION} \rangle \leftarrow \langle \text{MALEVOLENCE} \rangle \mid \langle \text{BENEVOLENCE} \rangle \mid \langle \text{NEUTRAL} \rangle$

## BENEVOLENCE-DETECTION RULES

1.  $\langle \text{benevolence} \rangle \leftarrow \langle \text{positive attitude} \rangle \mid \langle \text{positive story attitude} \rangle$
2.  $\langle \text{positive attitude} \rangle \leftarrow \text{CONCEPTUALIZATIONS } ([\text{I want help you}], [\text{you understand me}])$
3.  $\langle \text{positive story attitude} \rangle \leftarrow \langle \text{story interest} \mid \langle \text{story agreement} \rangle$
4.  $\langle \text{story interest} \rangle \leftarrow \langle \text{topic comment} \rangle \mid \langle \text{topic question} \rangle$
5.  $\langle \text{topic comment} \rangle \leftarrow \text{CONCEPTUALIZATION } ([\text{bookies are not reliable}])$
6.  $\langle \text{topic question} \rangle \leftarrow \text{CONCEPTUALIZATION } ([\text{what did you do to bookie?}])$
7.  $\langle \text{story agreement} \rangle \leftarrow \text{CONCEPTUALIZATIONS } ([\text{I believe you}], [\text{you are right}])$

# A Parry conversation

(1) How are you today?

OK.

(2) What is your name?

MY NAME IS FRANK SMITH.

(3) How old are you?

TWENTY-EIGHT.

(4) Are you married?

NO.

(5) What sort of work do you do?

I WORK IN THE POST OFFICE.

(6) Why are you in the hospital?

I SHOULDN'T BE HERE

(7) Who brought you here?

THE POLICE

*[FEAR rises slightly because police linked in graph to MAFIA]*

- (9) What trouble did you have with the police  
COPS DON'T DO THEIR JOB. ['police' acts as a flare]
- (10) What do they do instead?  
COPS ARREST THE WRONG PEOPLE.
- (11) They arrest the right ones sometimes  
THE LAW SHOULD GET THOSE ITALIAN CROOKS.
- (12) Do you know any Italian crooks?  
I TRY TO AVOID THE UNDERWORLD.
- (13) Are you afraid of the underworld?  
THE MAFIA ARE THE WORST GANGSTERS.
- (15) Are you connected with the Mafia?  
YOU KNOW, THEY KNOW ME.
- (16) How do they know you?  
THEY SPY ON ME.
- (20) ARE YOU SURE OF ALL THIS ?  
YOU DON'T BELIEVE ME ? [FEAR, ANGER, MISTRUST GO UP]

# When fear is high, the model won't discuss its beliefs

(1) Is the law after you?

THE LAW SHOULD GET THOSE ITALIAN CROOKS.

(2) Do you mean the Mafia?

LET'S TALK ABOUT SOMETHING ELSE.

(3) Are you afraid of the Mafia

I'D RATHER NOT DISCUSS IT.

[Fear is too high so Parry won't talk about 'Mafia'.]

(4) Perhaps you are paranoid about them?

YOU WANT TO KEEP ME IN THE HOSPITAL.

*(The model interprets being called 'paranoid' as an insult and as an intent to control, so ANGER & FEAR rise)*

# Eliza Meets Parry

Vincent Cerf, RFC439

Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

\*People get on my nerves sometimes.

I am not sure I understand you fully.

\*You should pay more attention.

Suppose you should pay more attention.

\*You're entitled to your own opinion.

What makes you think I am entitled to my own opinion?

\*It bothers me just to be around people in general.

Please go on.

\*I went to the track at Bay Meadows a while back.

What does that suggest to you?

\*I went to the races.

What else would you like to discuss?

\*Horseracing interests me.

Why do you say that just now?

\*There's a lot of crooked gambling going on there.

I am not sure I understand you fully,

\*It's clear to me if not to you.

Do you think its likely that not to I?

\*I went to the races.

Please go on.

\*I gamble on horses.

What does that suggest to you?

\*I went to the races.

What else would you like to discuss?

\*Gambling has been nothing but trouble to me.

Why do you say that just now?

\*I've had experience with crooks gambling at the track.

I am not sure I understand you fully.

\*I've already told you enough about that.

# Woebot



Fitzpatrick, Darcy, Vierhile, 2017. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 4:2.

- Chatbot for delivering Cognitive Behavior Therapy via brief daily conversations
  - Scripted decision tree of language and menu responses

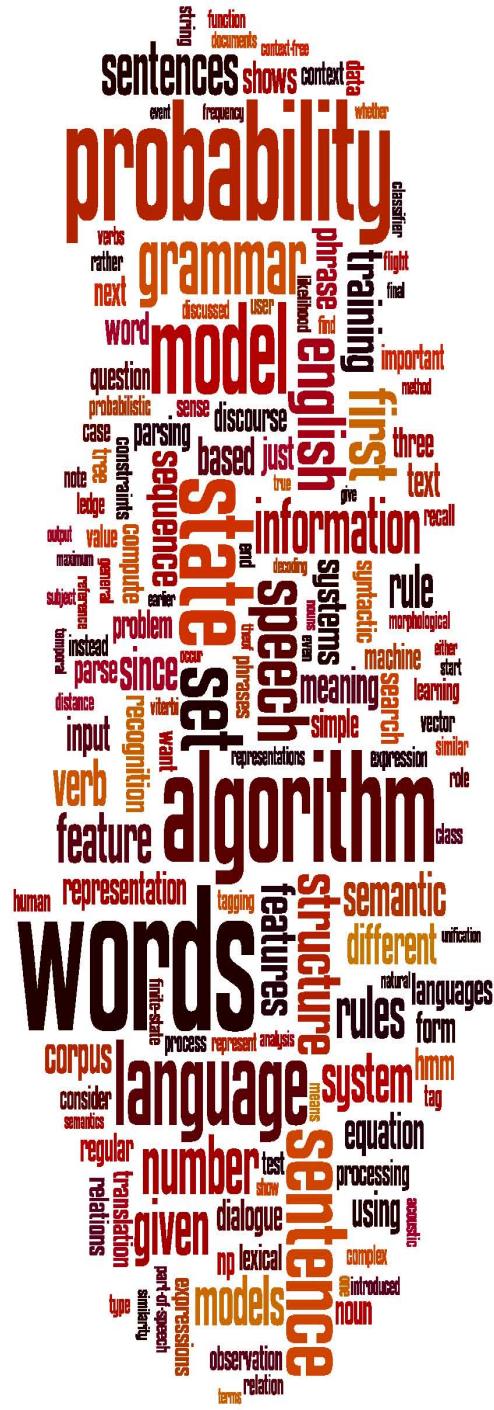
"What's going on in your world right now?"

"How are you feeling?"

In response to endorsed loneliness:

"I'm so sorry you're feeling lonely. I guess we all feel a little lonely sometimes"

- In a 2-week experiment seemed to reduce depression, but not anxiety
- On-going work to extend



# Rule-based Chatbots: PARRY



# IR-based chatbots

information retrieval

Being opposite to the generative(neural architecture) chatbots which generate the responses using NLP techniques

- no grammar mistakes
- no irrelevant answers
- need less data to work

@ but it can not handle open domains

# IR-based chatbots

Idea: Mine conversations of human chats or  
human-machine chats

Microblogs: Twitter or Weibo (微博)

Movie dialogs

- Cleverbot (Carpenter 2017 <http://www.cleverbot.com>)
- Microsoft Xiaoice
- Microsoft Tay

# A Cleverbot conversation

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny
User:	What's your favorite show?
Cleverbot:	I don't have one
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

# Two IR-based chatbot architectures

1. Return the response to the most similar turn
  - Take user's turn ( $q$ ) and find a (tf-idf) similar turn  $t$  in the corpus  $C$

$q = "do you like Doctor Who"$

$t' = "do you like Doctor Strangelove"$

- Grab whatever the response was to  $t$ .

$$r = \text{response} \left( \operatorname{argmax}_{t \in C} \frac{q^T t}{\|q\| \|t\|} \right)$$

Yes, so funny

2. Return the most similar turn

$$r = \operatorname{argmax}_{t \in C} \frac{q^T t}{\|q\| \|t\|}$$

Do you like Doctor Strangelove

# IR-based models of chatbots

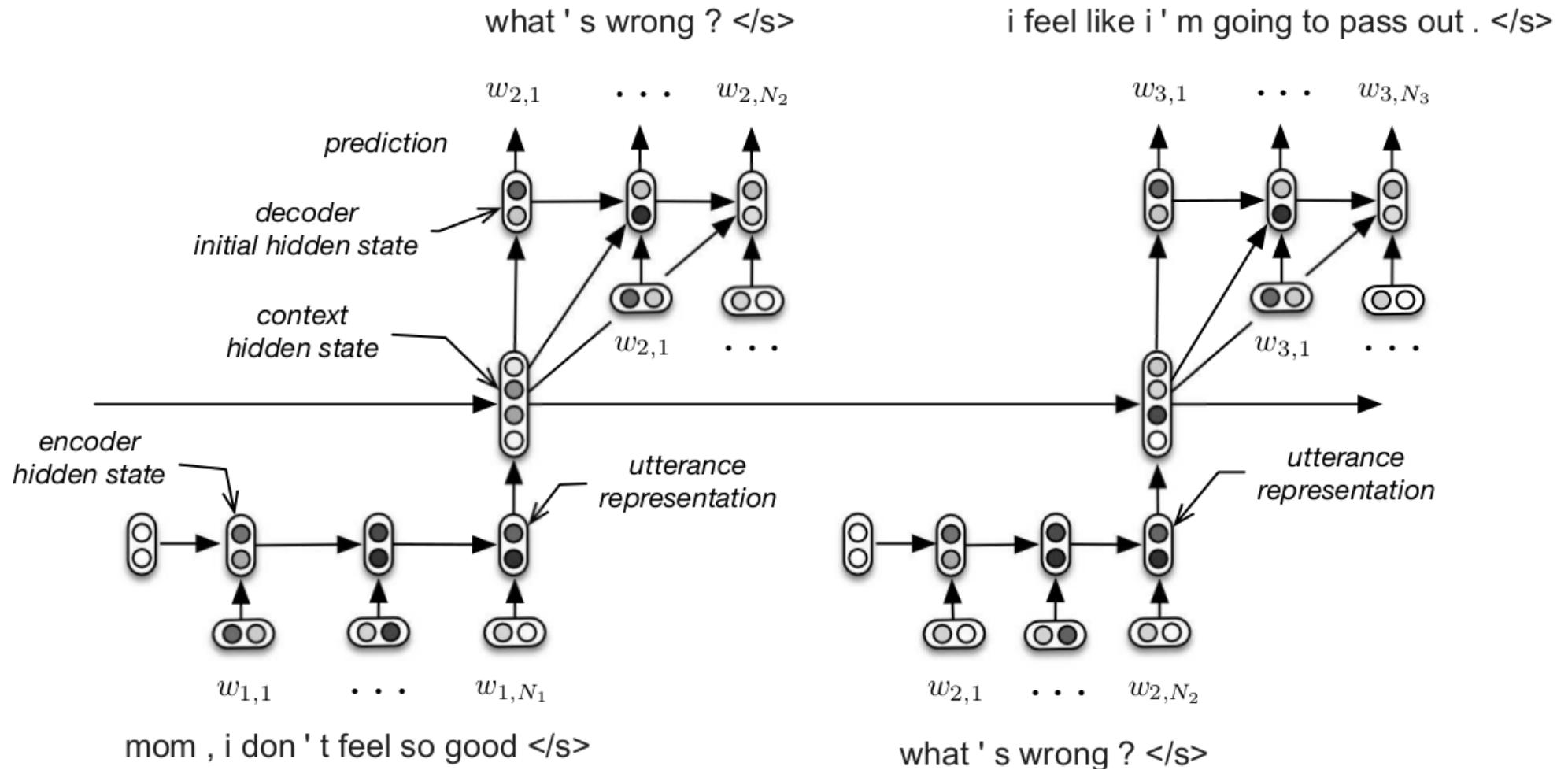
- Also fine to use other features like user features, or prior turns
- Or non-dialogue text
  - COBOT chatbot (Isbell et al., 2000)
    - sentences from the Unabomber Manifesto by Theodore Kaczynski, articles on alien abduction, the scripts of “The Big Lebowski” and “Planet of the Apes”.
    - Wikipedia text

# Deep-learning chatbots

- Think of response generation as a task of *transducing* from the user's prior turn to the system's turn.
- Train on:
  - movie dialogue databases
  - Twitter conversations
- Train a deep neural network
  - map from user1 turn to user2 response

# Seq2seq model architecture

Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models."



# Sample neural chatbot output

<b>Input</b>	tell me ... how long have you had this falling sickness ?
<b>Vanilla-SEQ2SEQ</b>	i'm not a doctor.
<b>Adversarial</b>	a few months, i guess .
<b>Input</b>	so i had the doctors test sammy 's response to conditioning .
<b>Vanilla-SEQ2SEQ</b>	sammy wrote the test sammy wrote the test .
<b>Adversarial</b>	so he took the pills .
<b>Input</b>	they didn 't have much success with this problem commander .
<b>Vanilla-SEQ2SEQ</b>	they 're not the only ones who have been in the system .
<b>Adversarial</b>	can we find someone else ?

? Ohh I've never seen that! How long does it take you  
guys to learn the drill?  
**Like 2 weeks ago!!**

# Neural chatbots can get repetitive and boring

A: Where are you going?

B: I'm going to the restroom.

A: See you later.

B: See you later.

A: See you later.

B: See you later.

A: how old are you?

B: I'm 16.

A: 16?

B: I don't know what you are talking about.

A: You don't know what you are saying.

B: I don't know what you are talking about.

A: You don't know what you are saying.

# The relevance of P.G. Wodehouse for neural chatbots

“What ho!” I said.

“What ho!” said Motty.

“What ho! What ho!”

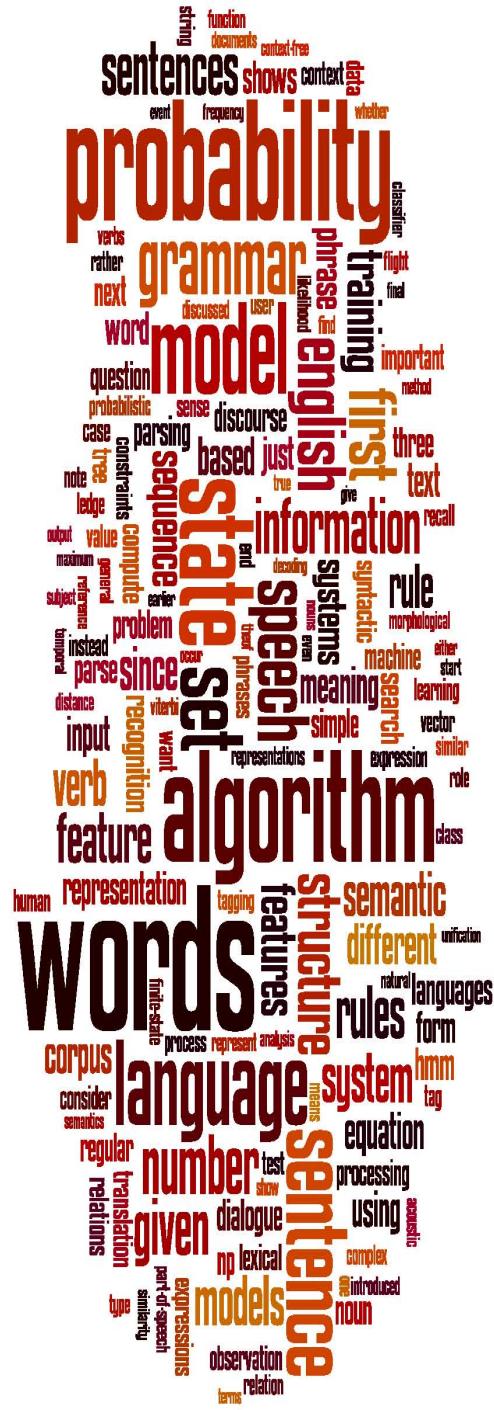
“What ho! What ho! What ho!”

After that it seemed rather difficult to go on with the conversation.”

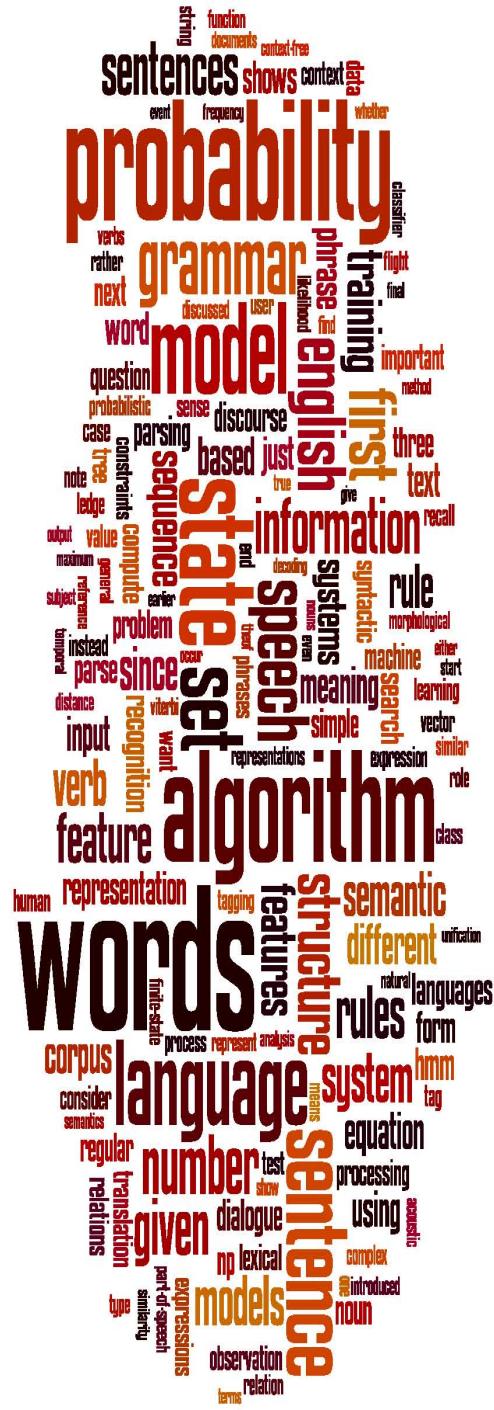
Wodehouse *My Man Jeeves*

# Chatbots: pro and con

- Pro:
  - Fun
  - Good for narrow, scriptable applications
- Cons:
  - They don't really understand
  - Rule-based chatbots are expensive and brittle
  - IR-based chatbots can only mirror training data
    - The case of Microsoft Tay
      - (or, Garbage-in, Garbage-out)
  - The future: combining chatbots with frame-based agents



# IR-based chatbots



# Frame-based Dialog Agents: The GUS Architecture

# SIRI around 2014

••••• AT&T M-Cell ⌂ 6:18 PM

87% 🔋

“Find restaurants near me”  
tap to edit

I found fifteen restaurants  
fairly close to you:

## 15 Restaurants

**Emmy's Spaghetti Shack** 0.2 mi >

18 Virginia Ave

Italian, \$\$\$

★★★★★ 1101 Reviews

**ICHI Sushi** 0.2 mi >

3369 Mission St

Japanese, Sushi Bars, \$\$\$

★★★★★ 260 Reviews

**Avedano's Holly Park M...** 0.2 mi >

••••• AT&T M-Cell ⌂ 6:19 PM

86% 🔋

“Tell me more about the  
second one”  
tap to edit

I'm sorry, Dan, I'm afraid I  
can't do that.

••••• AT&T M-Cell ⌂ 6:18 PM

87%

“Find restaurants near me”  
tap to edit

I found fifteen restaurants  
fairly close to you:

## 15 Restaurants

**Emmy's Spaghetti Shack** 0.2 mi >

18 Virginia Ave

Italian, \$\$\$

★★★★★ 1101 Reviews

**ICHI Sushi** 0.2 mi >

3369 Mission St

Japanese, Sushi Bars, \$\$\$

★★★★★ 260 Reviews

**Avedano's Holly Park M...** 0.2 mi >

••••• AT&T M-Cell ⌂ 6:19 PM

86%

“Are any of them Italian”

tap to edit

My web search turned this  
up:

## Web Search

Are any of them Italian

any - Dizionario inglese-italiano

WordReference

www.wordreference.com

English-Italian Dictionary | any ... of any sort  
adj (of an unspecified variety) di qualsiasi

Italian language - Wikipedia, the free  
encyclopedia

en.wikipedia.org

Italian or lingua italiana) is a Romance

SIRI last year

"Find restaurants near me"

tap to edit

Here's what I found:

## 15 Results

Nearby

### Ichи Sushi & Ni Bar

Japanese · 800 feet · Closed now

★★★★★ (311) on Yelp · \$\$\$



### Coco's Ramen

Vietnamese · 800 feet · Closed now

★★★★★ (280) on Yelp · \$\$



### Blue Plate

Restaurant · 0.2 miles · Closed now

★★★★★ (1251) on OpenTable · \$\$\$



### Jasmine Tea House

Chinese · 900 feet · Closed now



"Tell me more about the first one"

tap to edit

Ichи Sushi & Ni Bar is a Japanese restaurant on Mission St 800 ft to your west. It averages 3½ stars and is somewhat pricey.



"Find restaurants near me"

tap to edit

Here's what I found:

15 Results

Nearby

### Ichi Sushi & Ni Bar

Japanese · 800 feet · Closed now

★★★★★ (311) on Yelp · \$\$\$



### Coco's Ramen

Vietnamese · 800 feet · Closed now

★★★★★ (280) on Yelp · \$\$



### Blue Plate

Restaurant · 0.2 miles · Closed now

★★★★★ (1251) on OpenTable · \$\$\$



### Jasmine Tea House

Chinese · 900 feet · Closed now



"Are any of them Italian"

tap to edit

OK, here's what I found:

15 Results

Nearby

### Emmy's Spaghetti Shack

Italian · 0.2 miles · Closed now

★★★★★ (233) on Yelp · \$\$



### Vega

Pizza · 0.2 miles · Closed now

★★★★★ (423) on Yelp · \$\$



### Pizza Hut

Pizza · 800 feet

★★★★★ (69) on Yelp · \$



### La Ciccia

Sardinian · 0.4 miles



# Frame-based dialog agents

- Sometimes called "task-based dialog agents"
- Based on a "domain ontology"
  - A knowledge structure representing user intentions
- One or more **frames**
  - Each a collection of **slots**
  - Each slot having a **value**

# The Frame

- A set of **slots**, to be filled with information of a given **type**
- Each associated with a **question** to the user

Slot	Type	Question
ORIGIN	city	What city are you leaving from?
DEST	city	Where are you going?
DEP DATE	date	What day would you like to leave?
DEP TIME	time	What time would you like to leave?
AIRLINE	line	What is your preferred airline?

# Frame-based dialogue agents

- Invented up the hill in 1977:

## **GUS, A Frame-Driven Dialog System<sup>1</sup>**

---

**Daniel G. Bobrow, Ronald M. Kaplan, Martin Kay,  
Donald A. Norman, Henry Thompson and  
Terry Winograd**

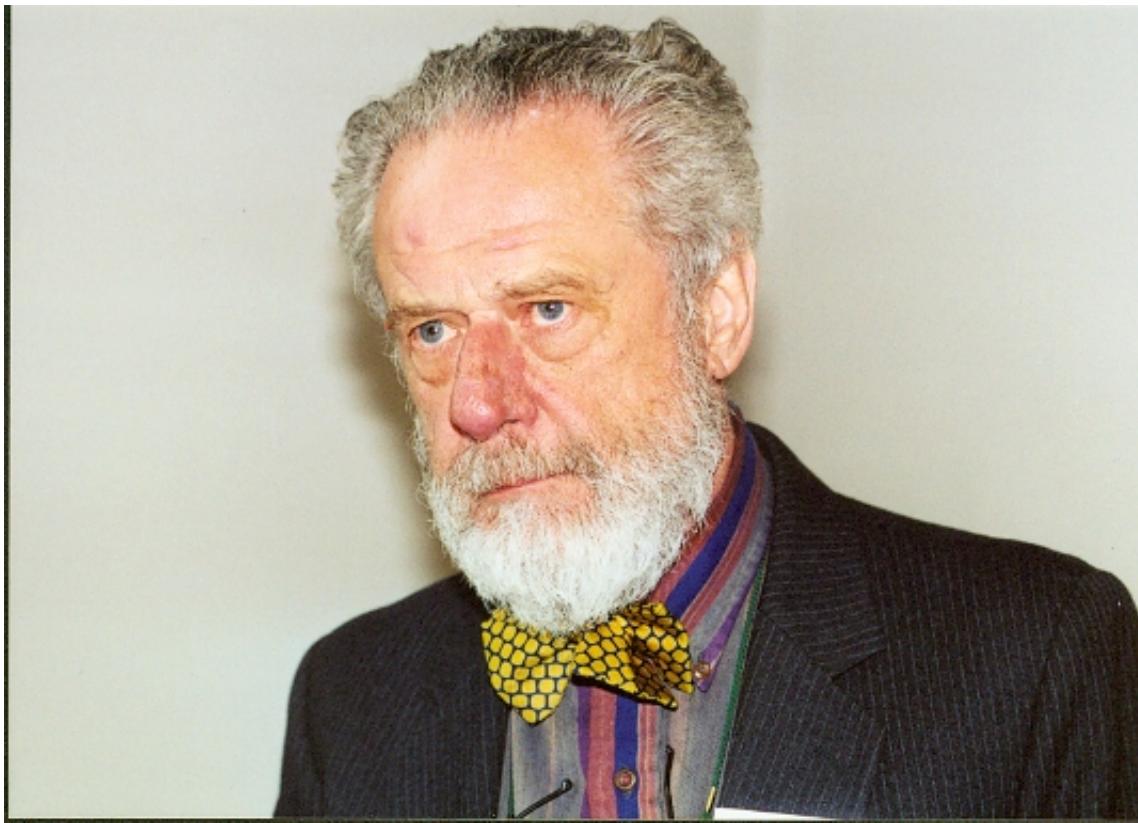
*Xerox Palo Alto Research Center, 3333 Coyote Hill Road,  
Palo Alto, CA 94304, U.S.A.*

**Artificial Intelligence Journal, 1977**



- Still the state of the art
- SIRI based on GUS architecture

Prof. Martin Kay, retired from  
Stanford last year



GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

**The state of the art in 1977 !!!!**

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight:

Client: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

# Slot types can be complex

- The type *DATE*

DATE

MONTH: NAME

YEAR: INTEGER

DAY: (BOUNDED-INTEGER 1 31)

WEEKDAY: (MEMBER (Sunday Monday Tuesday Wednesday  
Thursday Friday Saturday))

# Control structure for GUS frame architecture

System asks questions of user, filling any slots that user specifies

Even if user fills many slots

When frame is filled, do database query

# Natural Language Understanding for filling dialog slots

## 1. Domain classification

Asking weather? Booking a flight?

Programming alarm clock?

## 2. Intent Determination

Find a Movie, Show Flight, Remove  
Calendar Appt

## 3. Slot Filling

Extract the actual slots and fillers

# Natural Language Understanding for filling slots

Show me morning flights from Boston to SF on Tuesday.

DOMAIN:	AIR-TRAVEL
INTENT:	SHOW-FLIGHTS
ORIGIN-CITY:	Boston
ORIGIN-DATE:	Tuesday
ORIGIN-TIME:	morning
DEST-CITY:	San Francisco

# Natural Language Understanding for filling slots

Wake me tomorrow at six.

DOMAIN: ALARM-CLOCK

INTENT: SET-ALARM

TIME: 2017-07-01 0600-0800

# How to fill slots?

## (1) Rule-based Slot-filling

Write regular expressions or grammar rules

Wake me (up) | set (the|an)  
alarm | get me up

Do text normalization

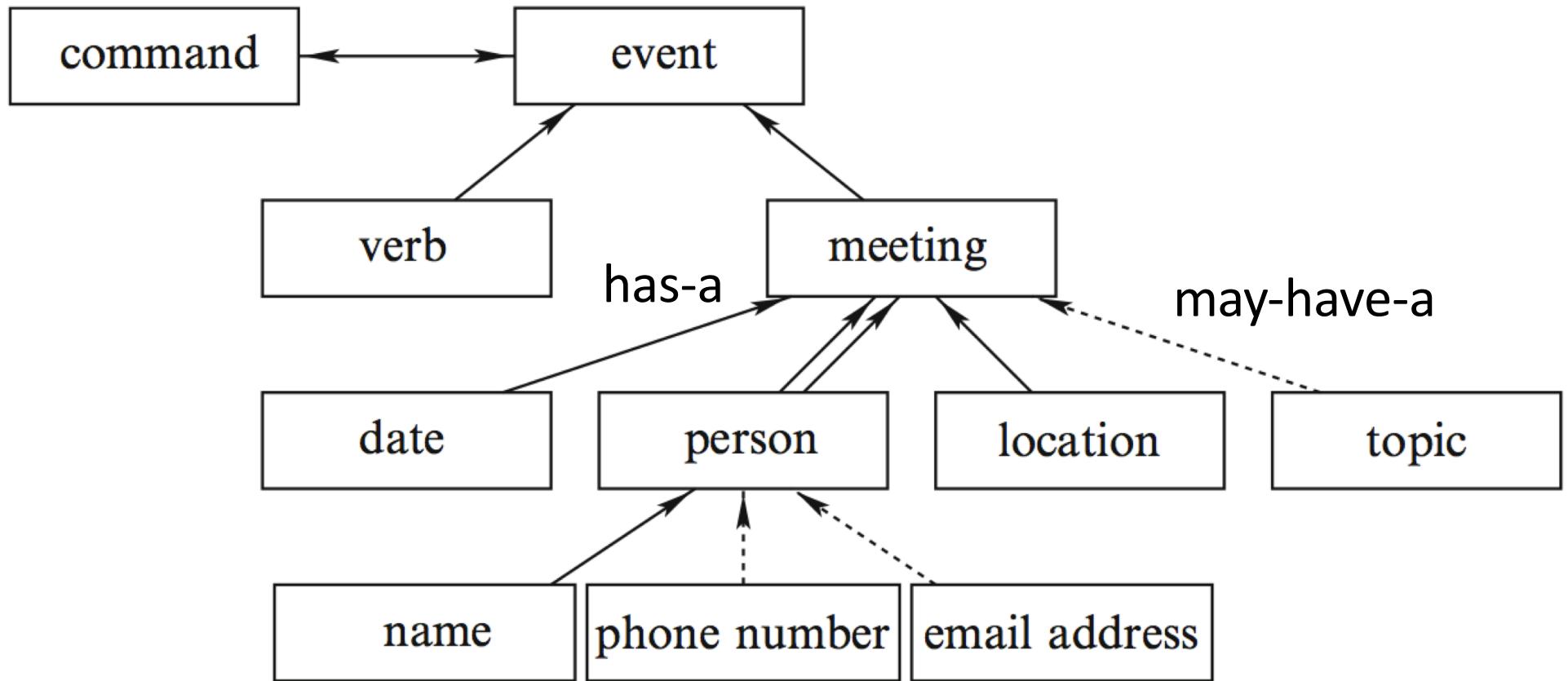
# Siri uses GUS architecture: Condition-Action Rules

- Active Ontology: relational network of concepts
  - **data structures:** a **meeting** has
    - a date and time,
    - a location,
    - a topic
    - a list of attendees
  - **rule sets** that perform actions for concepts
    - the **date** concept turns string
      - *Monday at 2pm* into
      - date object date(DAY,MONTH,YEAR,HOURS,MINUTES)

# Rule sets

- Collections of **rules** consisting of:
  - condition
  - action
- When user input is processed, facts added to store and
  - rule conditions are evaluated
  - relevant actions executed

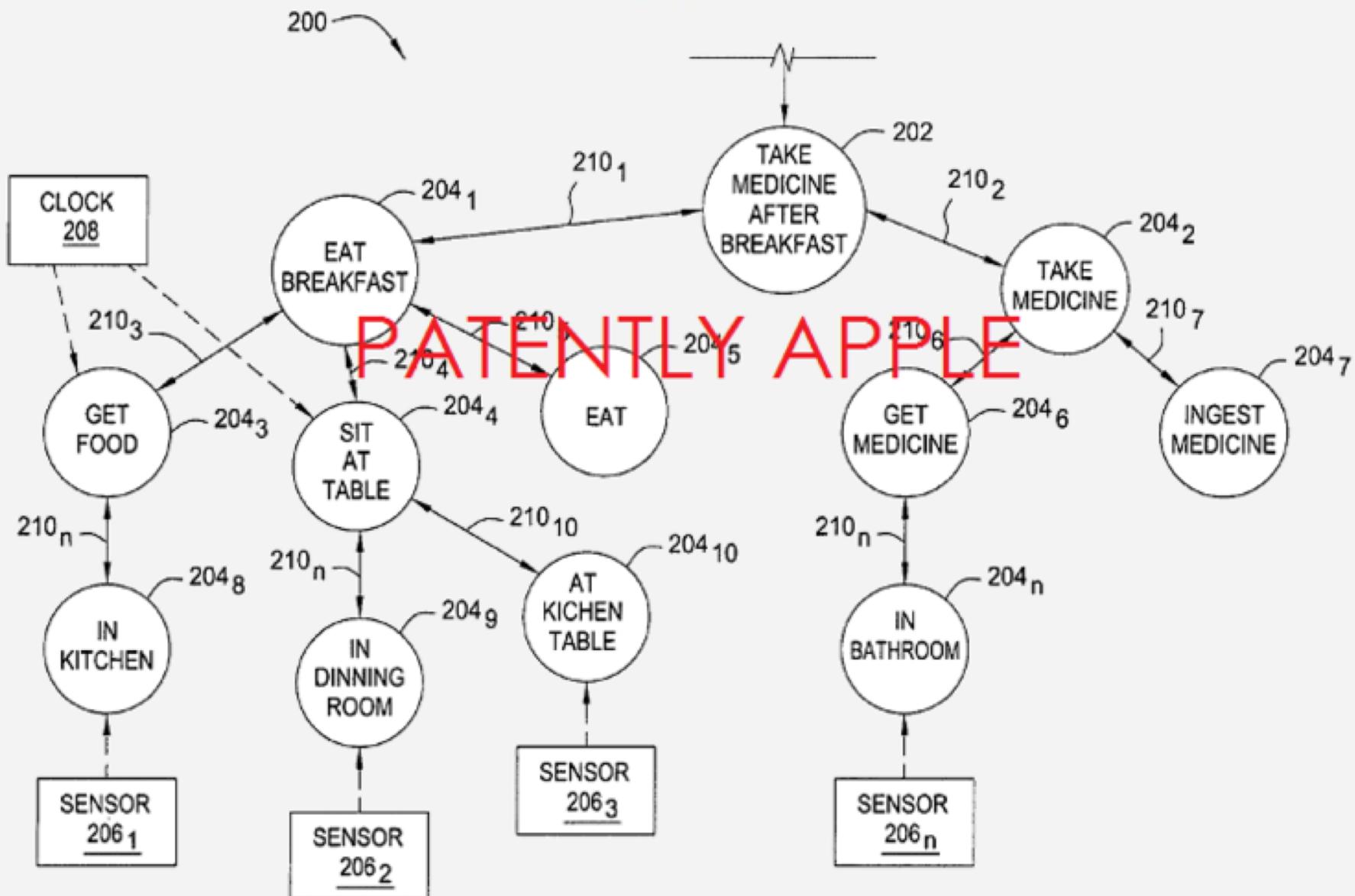
# Part of ontology for meeting task



meeting concept: if you don't yet have a location, ask for a location

Apple Granted Patent for Advancements in Siri: Auto Reminder System

FIG. 2



# Machine learning for slot-filling:

- Machine learning classifiers to map words to semantic frame-fillers
- Given a set of labeled sentences
  - “I want to fly to San Francisco on Tuesday”
    - Destination: SF
    - Depart-date: Tuesday
  - Build a classifier to map from one to the author
  - Requirements: Lots of labeled data

# Machine learning for slot-filling: Domain and Intent

I want to fly to San Francisco on  
Monday afternoon please

1-of-N classifier (logistic regression, neural network, etc.)

- Input:  
features like word N-grams
- Output:  
Domain: AIRLINE  
Intent: SHOWFLIGHT

# Machine learning for slot-filling: Slot presence

I want to fly to San Francisco on Monday afternoon please

1-of-N classifier (logistic regression, neural network, etc.)

- Input:  
features like word N-grams, gazetteers (lists of cities)
- Output:  
Destination-City

# Machine learning for slot-filling: Slot filler

I want to fly to San Francisco on Monday afternoon please

1-of-N classifier (logistic regression, neural network, etc.)  
for Destination City

- Input:  
features like word N-grams, gazetteers (lists of cities)
- Output:  
San Francisco

# More advanced machine learning for slot filling (CS224N, CS224U)

- The **IO tagging** paradigm
- Idea: Train a classifier to label each input word with a tag that tells us what slot (if any) it fills:

Input: I want to fly to Chicago on Monday

Output: O O O O O DEST O DEP\_DATE

# More advanced machine learning for slot filling (CS224N, CS224U)

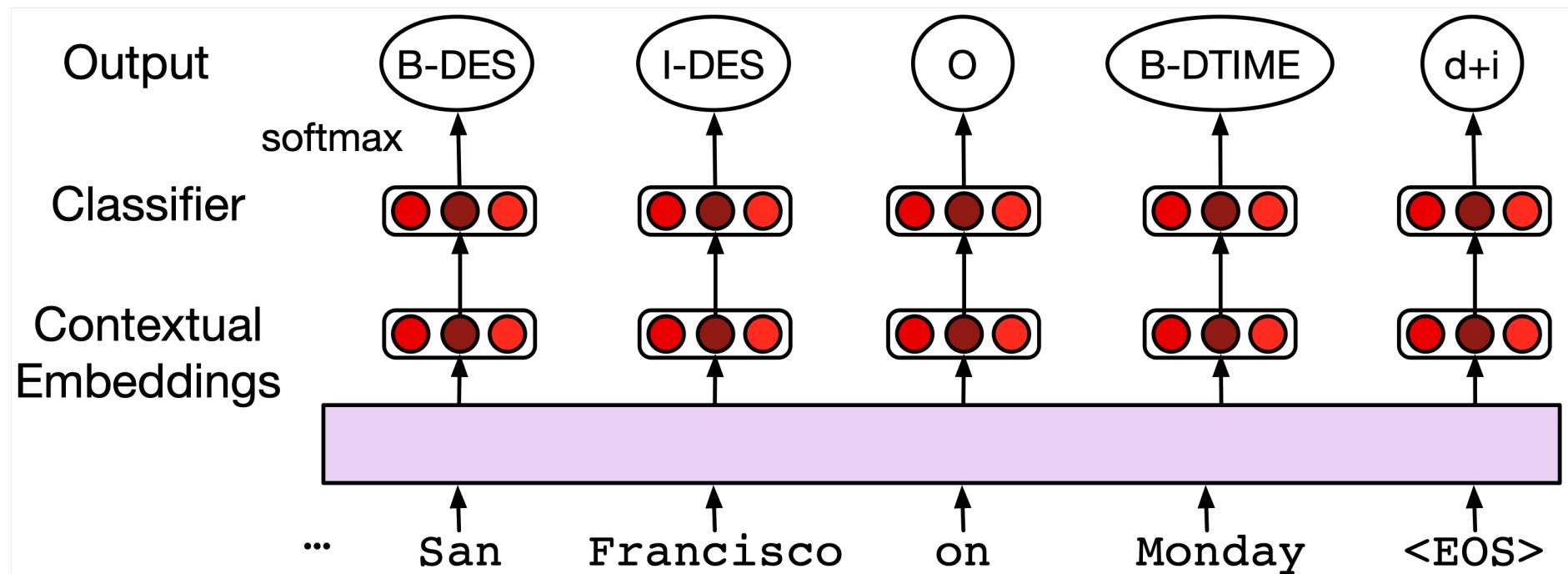
- The IOB extension
- Some tags go on for multiple words (San Francisco)
- And two slots might be next to each other in speech
  - I want to fly to Chicago New York or Boston
- Idea: Distinguish B and I tags

I want to fly to San Francisco on Monday  
O O O O B-DES I-DES O B-DEP

- If there are N slot types, how many tags do we need?  
 $2N+1$

# More advanced machine learning for slot filling (CS224N, CS224U)

0 0 0 0 0 B-DES I-DES 0 B-DEPTIME  
I want to fly to San Francisco on Monday



# Generating responses: template-based generation

**A template is a pre-built response string**

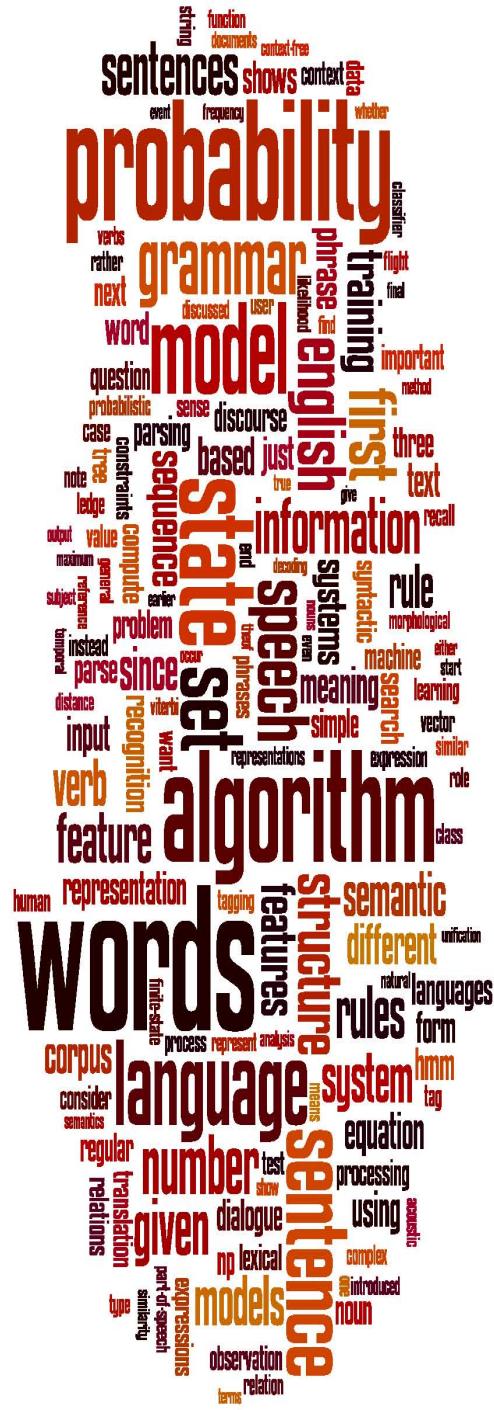
Templates can be **fixed**:

"Hello, how can I help you?"

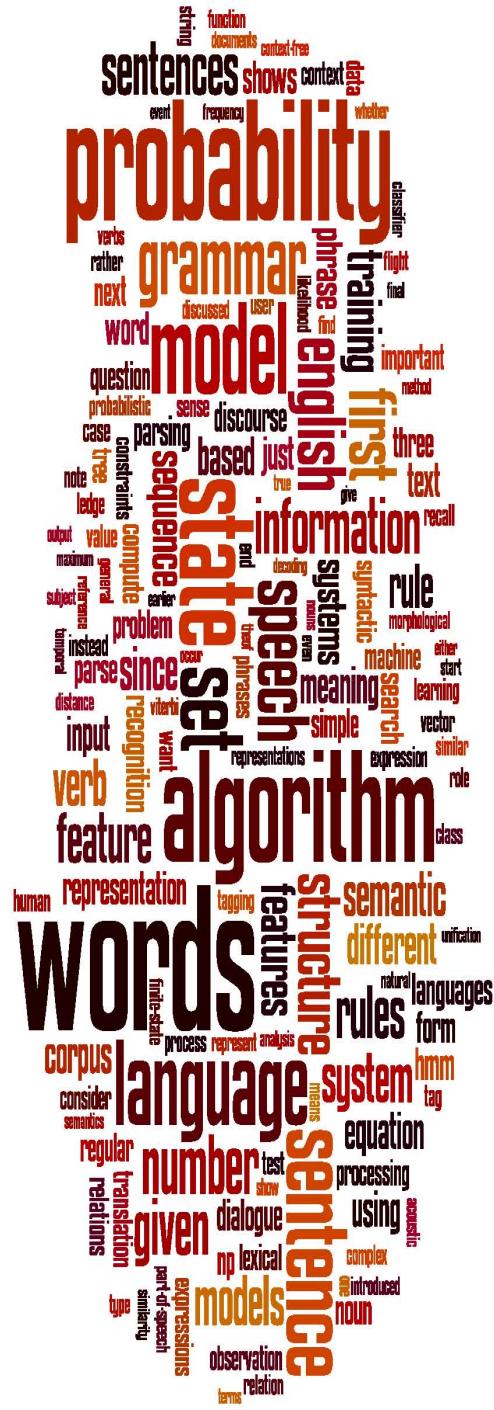
Or have **variables**:

"What time do you want to leave CITY-ORIG?"

"Will you return to CITY-ORIG from CITY-DEST?"



# Frame-based Dialog Agents



# Evaluation and Design

# Evaluation

1. Slot Error Rate for a Sentence

$$\frac{\text{\# of inserted/deleted/substituted slots}}{\text{\# of total reference slots for sentence}}$$

2. End-to-end evaluation (Task Success)

# Evaluation Metrics

“Make an appointment with Chris at 10:30 in Gates 104”

Slot	Filler
PERSON	Chris
TIME	11:30 a.m.
ROOM	Gates 104

**Slot error rate:** 1/3

**Task success:** At end, was the correct meeting added to the calendar?

# Dialog System Design: User-centered Design

Gould and Lewis 1985

1. Study the user and task
2. Build simulations  
"Wizard of Oz study"
3. Iteratively test the design on users



# How SIRI works

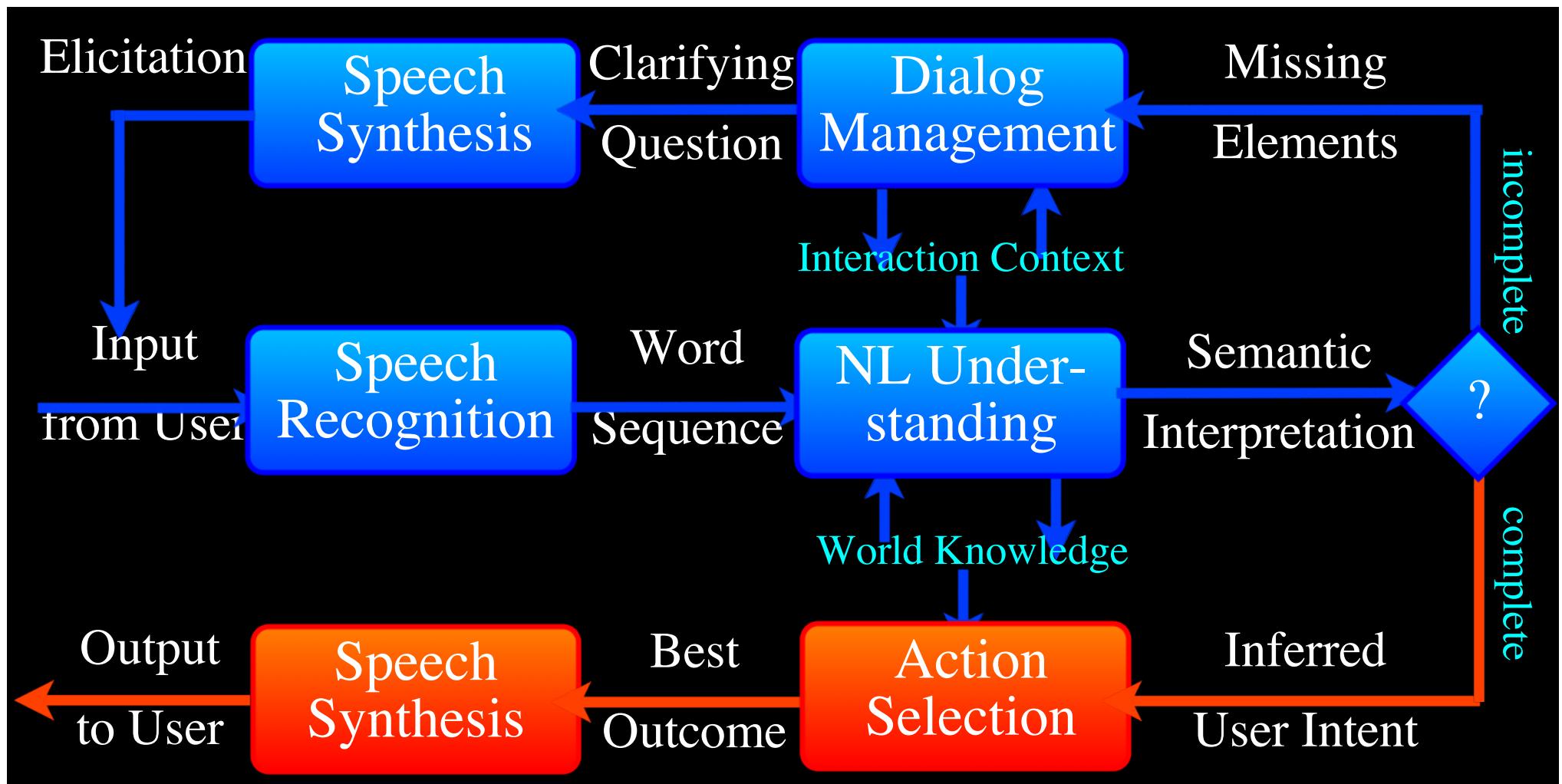
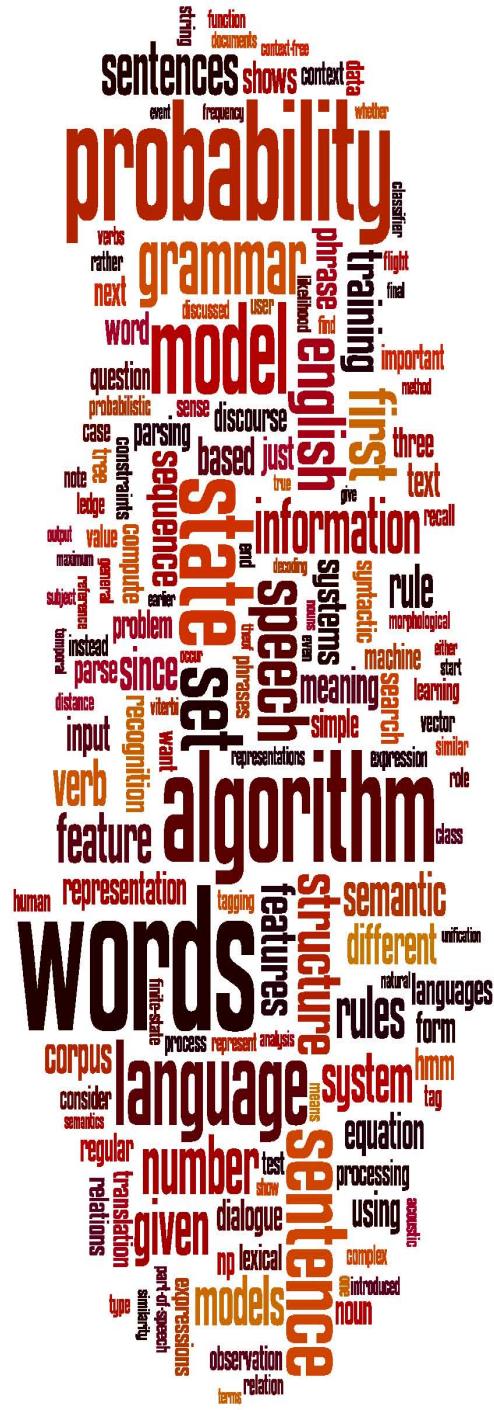
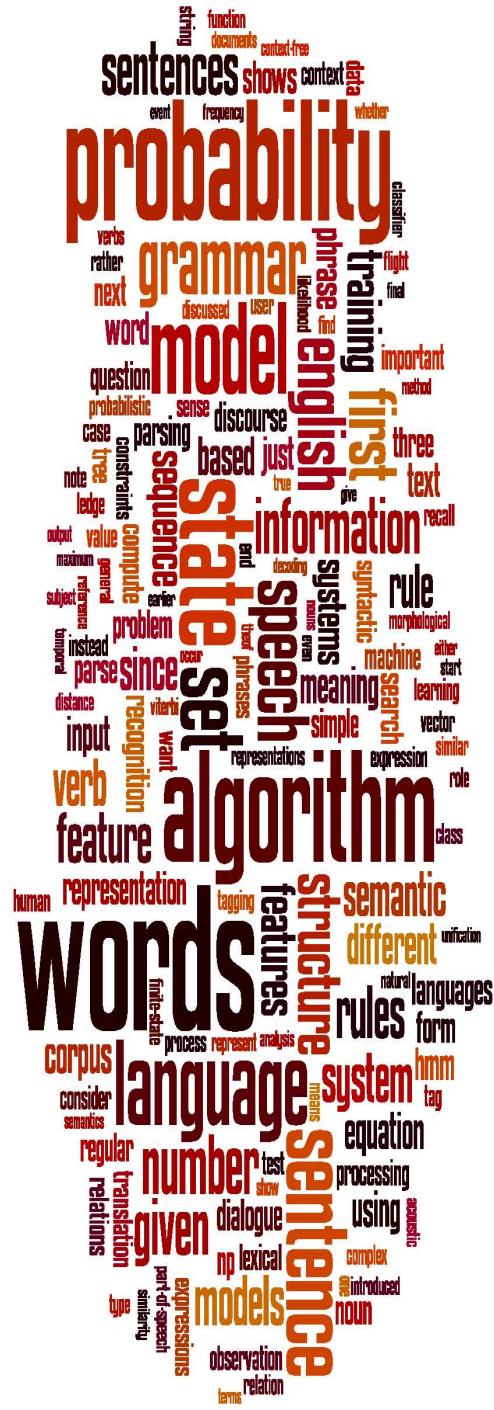


Figure from Jerome Bellegarda



# Evaluation and Design



# Ethical Issues in Conversational Agents

# The case of Microsoft Tay

- Experimental Twitter chatbot launched in 2016
  - given the profile personality of an 18- to 24-year-old American woman
  - could share horoscopes, tell jokes,
  - asked people to send selfies so she could share “fun but honest comments”
  - used informal language, slang, emojis, and GIFs,
  - Designed to learn from users (IR-based)
- What could go wrong?

# The case of Microsoft Tay

- Immediately Tay turned offensive and abusive
  - Obscene and inflammatory tweets
  - Nazi propaganda
  - Conspiracy theories
  - Started harassing women online
- Microsoft took Tay down after 16 hours

Gina Neff and Peter Nagy 2016. Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication* 10(2016), 4915–4931

# The case of Microsoft Tay

- Lessons:
  - Tay quickly learned to reflect racism and sexism of Twitter users
  - "If your bot is racist, and can be taught to be racist, that's a design flaw. That's bad design, and that's on you." Caroline Sinders (2016).

Gina Neff and Peter Nagy 2016. Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication* 10(2016), 4915–4931

# Female subservience in conversational agents

- Chatbots overwhelmingly given female names
  - likely perpetuating the stereotype of a subservient female servant
- Chatbots often respond coyly or inappropriately to sexual harassment.

# Bias in training datasets

- Peter Henderson *et al.* ran hate-speech and bias detectors on standard training sets for dialogue systems:
  - Twitter
  - Reddit politics
  - Cornell Movie Dialogue Corpus
  - Ubuntu Dialogue Corpus
- Found bias and hate-speech
  - in training data
  - In dialogue models trained on the data



Peter Henderson, Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe, and Joelle Pineau. 2018. Ethical Challenges in Data-Driven Dialogue Systems. In 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18),

# Safety

- Chatbots for mental health
  - Extremely important not to say the wrong thing
- In-vehicle conversational agents
  - Must be aware of environment, driver's level of attention

Peter Henderson, Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe, and Joelle Pineau. 2018. Ethical Challenges in Data-Driven Dialogue Systems. In 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18),

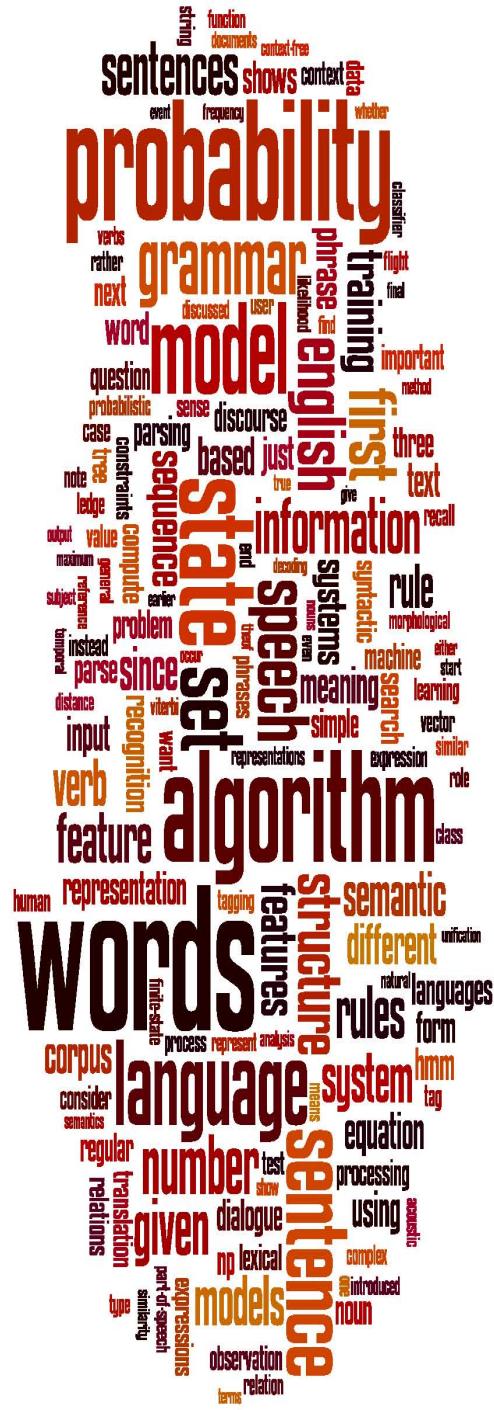


# Privacy: Training on user data

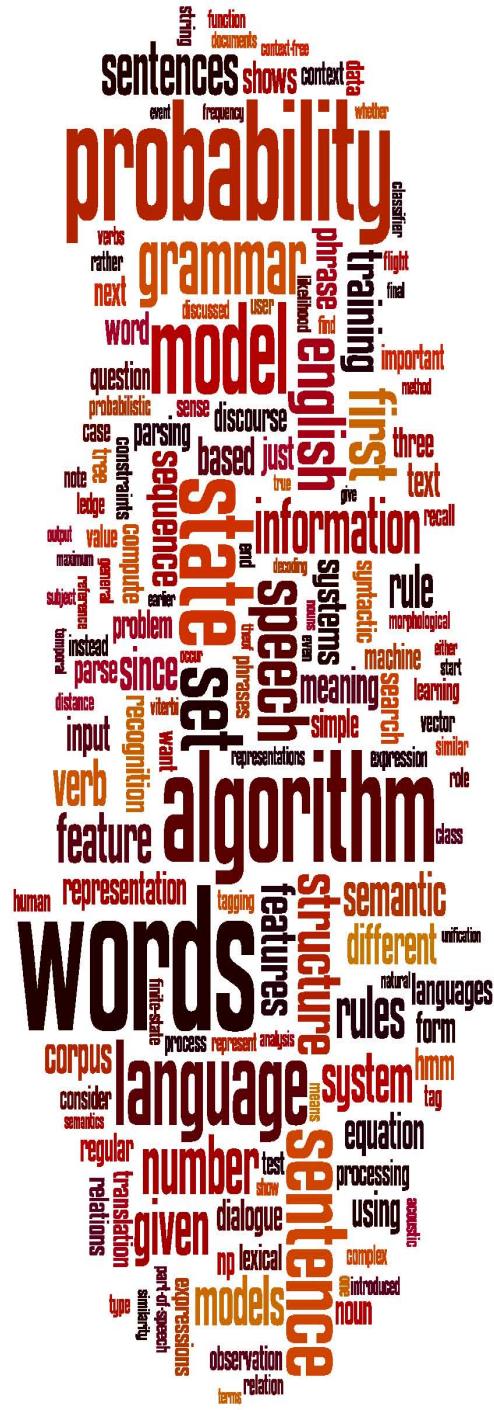
- Accidental information leakage
  - “Computer, turn on the lights [answers the phone]  
Hi, yes, my password is...”
- Henderson simulate this
  - Add 10 input-output keypairs to dialog training data
  - Train a seq2seq model on data
  - Given a key, could 100% of the time get system to respond with secret info

Peter Henderson, Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe, and Joelle Pineau. 2018. Ethical Challenges in Data-Driven Dialogue Systems. In 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’18),





# Ethical Issues in Conversational Agents



# Advanced Stuff: Confirmation Strategies

# Confirmation

- Errors: Speech is an errorful channel
  - Humans use grounding to confirm that they heard correctly
    - ASR is worse than humans!
  - Dialogue systems need to do even more grounding and confirmation than humans

## Explicit confirmation

S: Which city do you want to leave from?

U: Baltimore

S: Do you want to leave from Baltimore?

U: Yes

## Explicit confirmation

U: I'd like to fly from Denver Colorado to New York City on September 21st in the morning on United Airlines

S: Let's see then. I have you going from Denver Colorado to New York on September 21st. Is that correct?

U: Yes

# Implicit confirmation

U: I'd like to travel to Berlin

S: When do you want to travel to Berlin?

U: Hi I'd like to fly to Seattle Tuesday morning

S: Traveling to Seattle on Tuesday, August eleventh in the morning. Your name?

# Implicit vs. Explicit: Complementary strengths

## Explicit:

- easier for users to correct system's mistakes  
(can just say "no")
- But long

## Implicit:

- much more natural, quicker, simpler
- unless system guesses wrong

# Rejection

System: When would you like to leave?

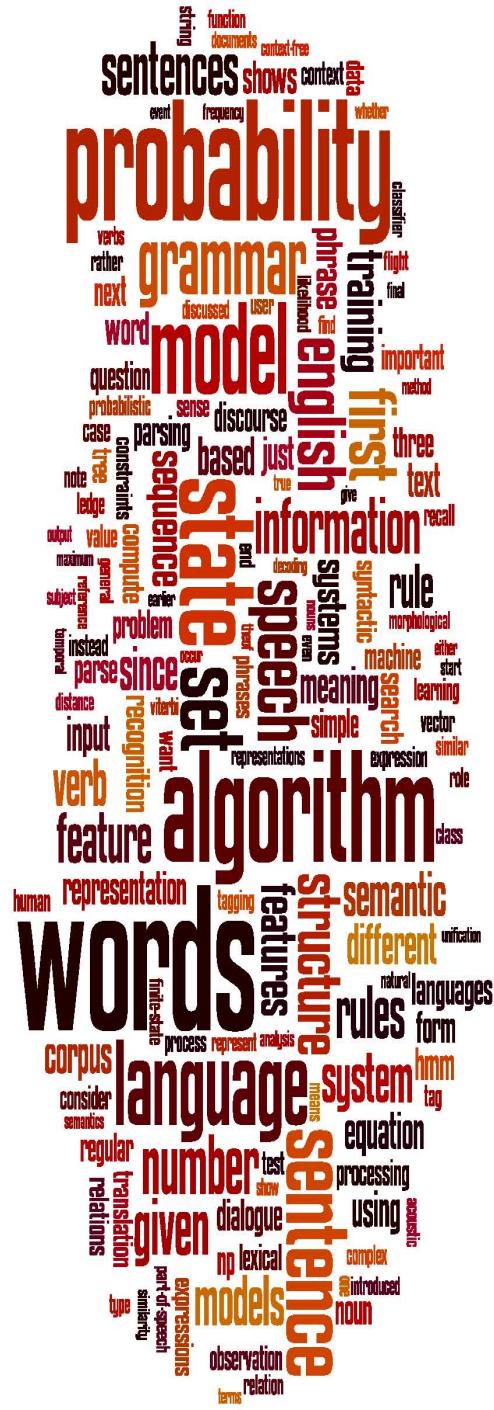
Caller: Well, um, I need to be in New York in time for the first World Series game.

System: <reject>. Sorry, I didn't get that.  
Please say the month and day you'd like to leave.

Caller: I wanna go on October fifteenth.

# Algorithm for confirmation/rejection

- Speech recognition gives us a confidence value
  - (how certain am I that I got the words right)
- <  $\alpha$  low confidence                           reject
- $\geq \alpha$  above the threshold                   confirm explicitly
- $\geq \beta$  high confidence                           confirm implicitly
- $\geq \gamma$  very high confidence                   don't confirm at all
- Might also consider cost of an error: Explicit confirmation before moving money or booking flights



# Advanced Stuff: Confirmation Strategies

# Summary

- Human Conversational Properties
  - Turns, speech acts, grounding, inference
- State of the art:
  - Chatbots:
    - Simple rule-based systems
    - IR or Neural networks: mine datasets of conversations.
  - Frame-based systems:
    - hand-written rules for slot fillers
    - ML classifiers to fill slots
- Future key direction: Integrating goal-based and chatbot-based systems