

計算言語学₁₁

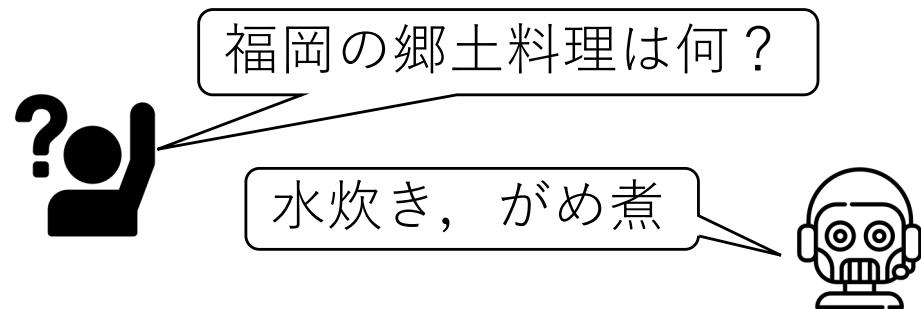
言語処理応用
(質問応答と対話)

東京大学生産技術研究所
吉永 直樹

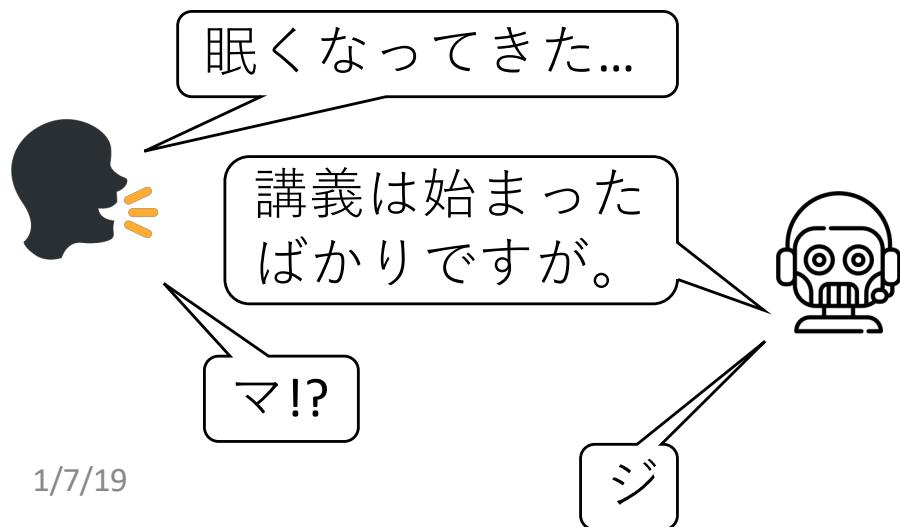
site: <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/class/cl/>

(エンドユーザを対象とした)言語処理応用

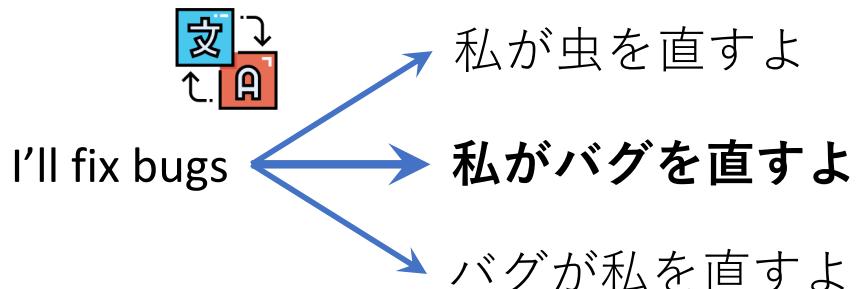
質問応答



対話システム

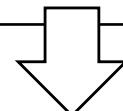


機械翻訳



自動要約

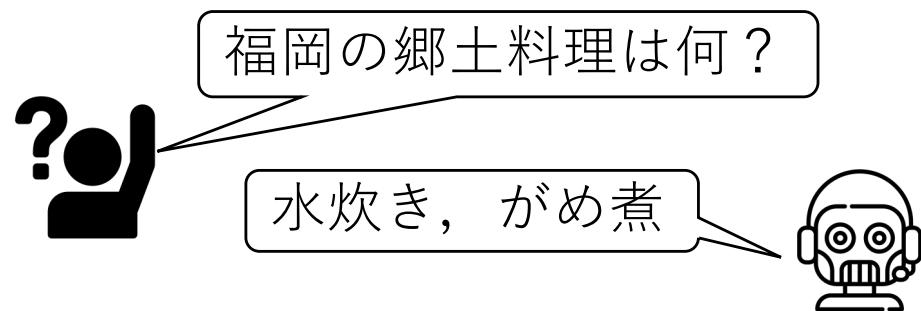
John Forbes Nash Jr. (June 13, 1928 – May 23, 2015) was an American mathematician who made fundamental contributions to game theory, differential geometry, and the study of partial differential equations.^{[2][3]} Nash's work has provided insight into the factors that govern chance and decision-making inside complex systems found in everyday life. His theories are widely used in economics. Serving as a Senior Research Mathematician at Princeton University during the later part of his life, he shared the 1994 Nobel Memorial Prize in Economic Sciences with game theorists Reinhard Selten and John Harsanyi. In 2015, he also shared the Abel Prize with Louis Nirenberg for his work on nonlinear partial differential equations. John Nash is the only person to be awarded both the Nobel Memorial Prize in Economic Sciences and the Abel Prize. In 1959, Nash began showing clear signs of mental illness, and spent several years at psychiatric hospitals being treated for paranoid schizophrenia. After 1970, his condition slowly improved, allowing him to return to academic work by the mid-1980s.^[4] His struggles with his illness and his recovery became the basis for Sylvia Nasar's biography, *A Beautiful Mind*, as well as a film of the same name starring Russell Crowe as Nash.^{[5][6][7]} On May 23, 2015, Nash and his wife Alicia were killed in a car crash while riding in a taxi on the New Jersey Turnpike.



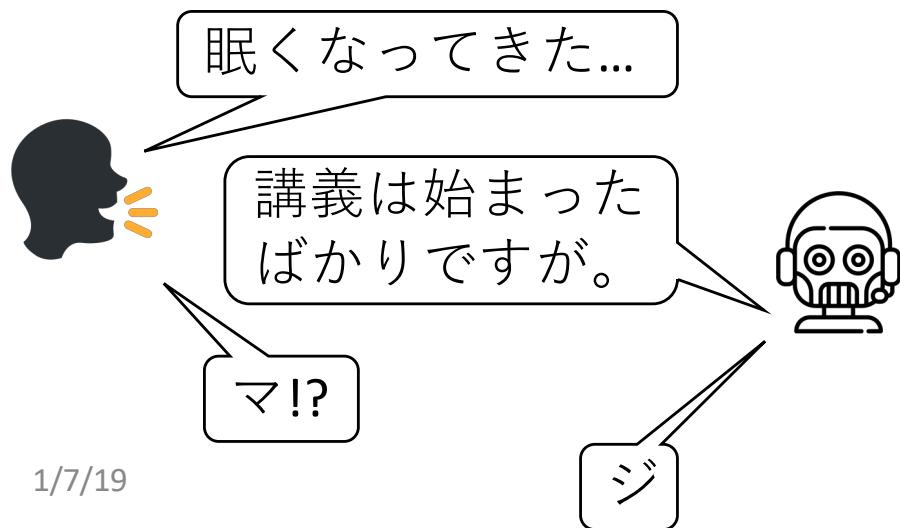
John Nash is a mathematical genius.

(エンドユーザを対象とした)言語処理応用

質問応答

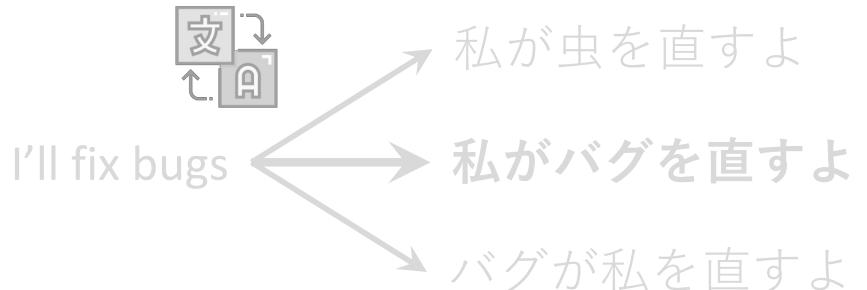


対話システム



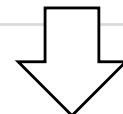
1/7/19

機械翻訳



自動要約

John Forbes Nash Jr. (June 13, 1928 – May 23, 2015) was an American mathematician who made fundamental contributions to game theory, differential geometry, and the study of partial differential equations.^{[1][2]} Nash's work has provided insight into the factors that govern chance and decision-making inside complex systems found in everyday life. His theories are widely used in economics. Serving as a Senior Research Mathematician at Princeton University during the later part of his life, he shared the 1994 Nobel Memorial Prize in Economic Sciences with game theorists Reinhard Selten and John Harsanyi. In 2015, he also shared the Abel Prize with Louis Nirenberg for his work on nonlinear partial differential equations. John Nash is the only person to be awarded both the Nobel Memorial Prize in Economic Sciences and the Abel Prize. In 1959, Nash began showing clear signs of mental illness, and spent several years at psychiatric hospitals being treated for paranoid schizophrenia. After 1970, his condition slowly improved, allowing him to return to academic work by the mid-1980s.^[3] His struggles with his illness and his recovery became the basis for Sylvia Nasar's biography, *A Beautiful Mind*, as well as a film of the same name starring Russell Crowe as Nash.^{[4][5][6][7]} On May 23, 2015, Nash and his wife Alicia were killed in a car crash while riding in a taxi on the New Jersey Turnpike.



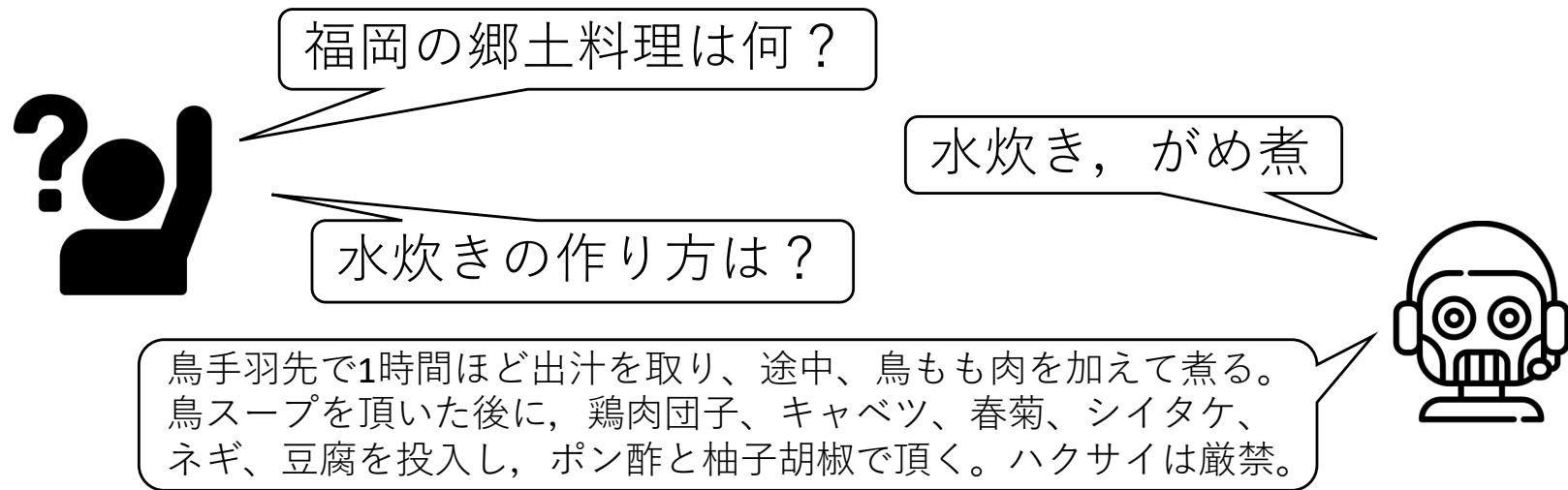
John Nash was a mathematical genius. 3

質問応答システム

Question Answering System

質問応答システム

- 自然言語の質問文に対し、自然言語で回答を行う



- TREC-8 (1999) で評価タスクが設定され、急速に発展
 - Factoid 型質問に限定すれば人に勝る性能 (IBM Watson)
- 近年は(ニューラル)機械読解のベンチマークとして主に研究が行われている

質問応答システム: サポートする質問タイプ

- ファクト型質問
事実を単語や句で解答する
タイプの質問
 - *What currency is used in Japan?*
-- the yen
 - *Where is the Skellig Michael?*
-- county Kerry, Ireland
 - *When did T. Cook become CEO?*
-- August, 2011

- ノンファクト型質問
定義・理由・方法などを文
で解答するタイプの質問
 - *Why can birds fly?*
-- Because their wings create
airfoils that can split the air, and
the difference between ...
 - *How to solve aging population?*
-- A long-term healthcare
coverage for various services will
make ...

ノンファクト型の質問応答システムは未だ発展途上

質問応答システム: 基本的なアプローチ

- Knowledge-based
 - [Green+ 1961]
 - 知識源: 知識ベース (DB)
 - 手法: 意味表現の照合
 - 出力: DB の該当エントリ
- Information Retrieval (IR)-based
 - 知識源: テキスト (Web等)
 - 手法: 情報検索 + α
 - 出力: 質問文の回答となる文書断片

知識に基づく質問応答システム (1/2)

- 知識ベース (DB) を知識源とし、質問をDBのクエリ (意味表現) に変換し該当するDBのエントリを取得
 - 知識ベース: 関係DB, **RDF triple** (semantic web)

When was John Lennon born?
→ birth-year (John Lennon, ?x)

Subject	Predicate	object
John Lennon	birth-year	1940

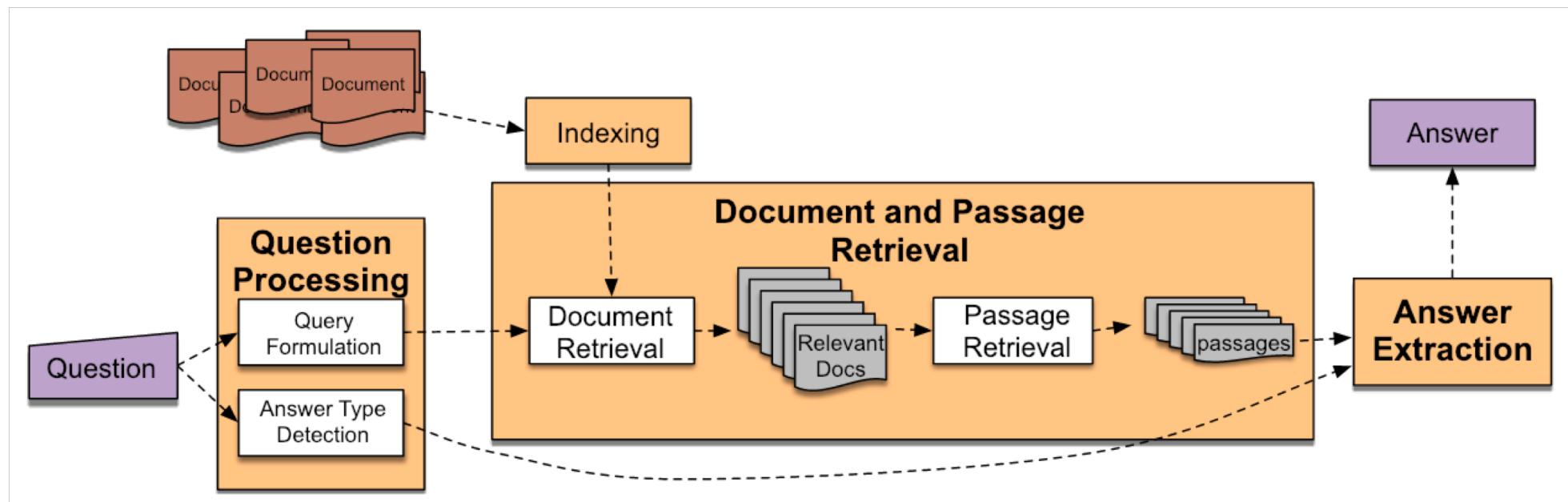
- ルールや教師あり学習に基づく手法
cf. <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/class/cl/cl10.pdf> p.43-
質問文の表現の多様性への対処が困難な場合も
- 半教師あり学習に基づく手法
 - **Distant supervision, Open IE, 言い換え表現の利用など**

知識に基づく質問応答システム (2/2): 質問文の表現の多様性への対処

- **Distant Supervision:** 大規模テキスト中に知識ベースのエントリが出現する箇所を正例とし分類器を学習
- **Open IE (Information Extraction)** で文から抽出した triple (subj-relation-obj) と知識ベースの対応を学習
 - Entity Linking で subj/obj を知識ベースのエントリに割当
- 言い換え表現を利用した質問文の換言
 - WikiAnswers でユーザが同義とタグ付けした質問文から言い換え表現を収集 [Fader+ 2013]

情報検索に基づくファクト型質問応答

- (大規模)テキストを知識源として、**情報検索 (IR)** を用いて回答が書かれている箇所を収集



情報検索に基づくファクト型質問応答: 質問文解析 (1/3)

- **検索クエリ生成**: 質問文を解析し知識源文書の検索のためのクエリ(検索キーワード)を生成
- 解答抽出のための補足情報も抽出
 - 解答タイプ: 解答となる固有表現の種別(人/場所/時間等)
 - フォーカス: 解答で置換(具体化)される文字列
 - 質問タイプ: 質問の種別(定義, 計算, リスト等)

例

Which US state capital has the largest population?

- クエリ: US state capital has the largest population
- 解答タイプ: 都市
- フォーカス: state, capital
- 質問タイプ: definition

情報検索に基づくファクト型質問応答: 質問文解析 (2/3)

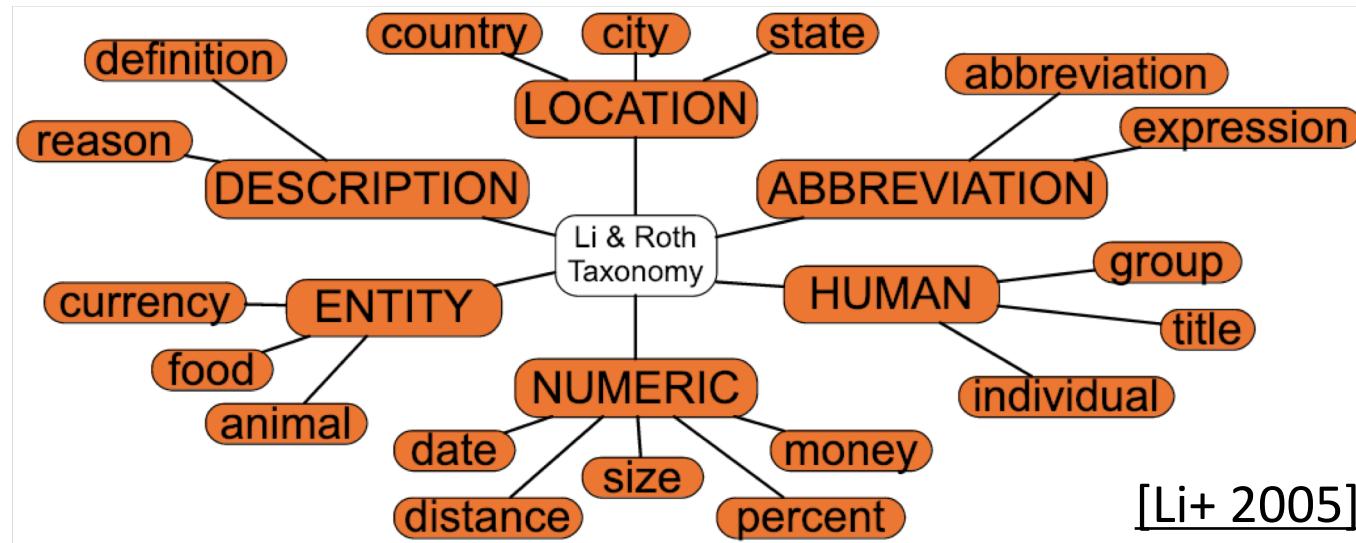
- **検索クエリ生成**: 質問文を解析し知識源文書の検索のためのクエリ(検索キーワード)を生成
 - Web を知識源とする場合、単に疑問詞を除けば良い
例) *Which US state capital has the largest population?*
 - 小規模な知識源を用いる場合、query expansion で適合率を改善したり、検索時の tf-idf を句で計算して精度改善
[Chen+ 2017]
- ルールに基づくクエリ再構成 (query reformulation)
[Lin 2007] により、検索し易くする場合も

When was the laser invented?
→ *the laser was invented*

where is the Valley of the Kings?
→ *the Valley of the Kings is located*

情報検索に基づくファクトトイド型質問応答: 質問文解析 (3/3)

- 解答タイプ(回答となるエンティティの種類)の認識
 - 固有表現 or Wikipedia から構築した taxonomy など



- パターン[Hovy+ 2002]や教師あり学習でタイプを類推
 - 特徴量: 質問文中の単語(埋め込み, 品詞, 固有名詞)

情報検索に基づくファクト型質問応答: 文書・メッセージ検索

- 検索エンジンでクエリを含む $\text{top-}n$ 文書を取得後、解答を含む可能性が高い **メッセージ** を抽出
 - メッセージ (passage): セクション, 段落, 文など
 - Web 検索を用いる場合, **スニペット** を利用可
- メッセージ検索では以下を考慮し, メッセージの絞り込みまたはランキングを行うことも
 - 正解の回答タイプの固有表現の数
 - 含まれるクエリ語の数
 - 含まれるクエリ語中の最長 n -gram
 - 文書検索における検索順位
 - クエリ語間の近接度 [Pasca 2003, Monz 2004]
 - 質問文との n -gram のオーバーラップ [Brill+ 2002]

情報検索に基づくファクトトイド型質問応答: 解答抽出

- タスク: 与えられたパッセージから解答となる
箇所(スパン)を認識

How tall is Mt. Everest?

Passage: *The official height of Mount Everest is 29029 feet*

- N-gram tilling: Web の冗長性を利用 [Brill+ 2002, Lin 2007]
 - スニペット中の各 n-gram を(重み付き)文書頻度や回答タイプとの合致度に応じスコア付けし重複を考慮して結合
- 教師あり学習に基づく手法
 - 人手で設計した特徴量を利用 [Pasca 2003]
 - 深層学習に基づく手法 [Chen+ 2017]

情報検索に基づくファクトトイド型質問応答: 人手で設計した特徴量に基づく教師あり解答抽出

- 与えられた文またはスパンが回答を含むかどうか
判定する分類器をラベル付きデータから学習
- 特徴量:
 - 回答候補が正解の回答タイプの句を含むかどうか
 - 回答候補にマッチする回答パターンがあるかどうか
 - 回答候補に含まれるクエリ語の数
 - 回答候補とクエリの距離
 - 回答候補がクエリに含まれない語を含むかどうか
 - 回答候補がクエリ語を多く含む語の同格表現であるか
 - 回答候補に後続する語が句読点かどうか
 - 回答候補中に含まれるクエリ中の最長 n-gram の長さ

情報検索に基づくファクト型質問応答: ニューラル解答抽出 (1/2)

- 入力クエリ(のベクトル表現)と意味的類似度の高い(ベクトル表現を持つ)句を解答として抽出
 - Machine Reading (機械読解)のベンチマークタスクとして独立に研究が行われている
 - 注) 既存のデータセットは、文書読解のベンチマークとしては問題あり [Sugawara+ 2017, 2018]

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

Q: "In what city and state did Beyoncé grow up?"

A: "Houston, Texas"

Q: "What areas did Beyoncé compete in when she was growing up?"

A: "singing and dancing"

Q: "When did Beyoncé release Dangerously in Love?"

A: "2003"

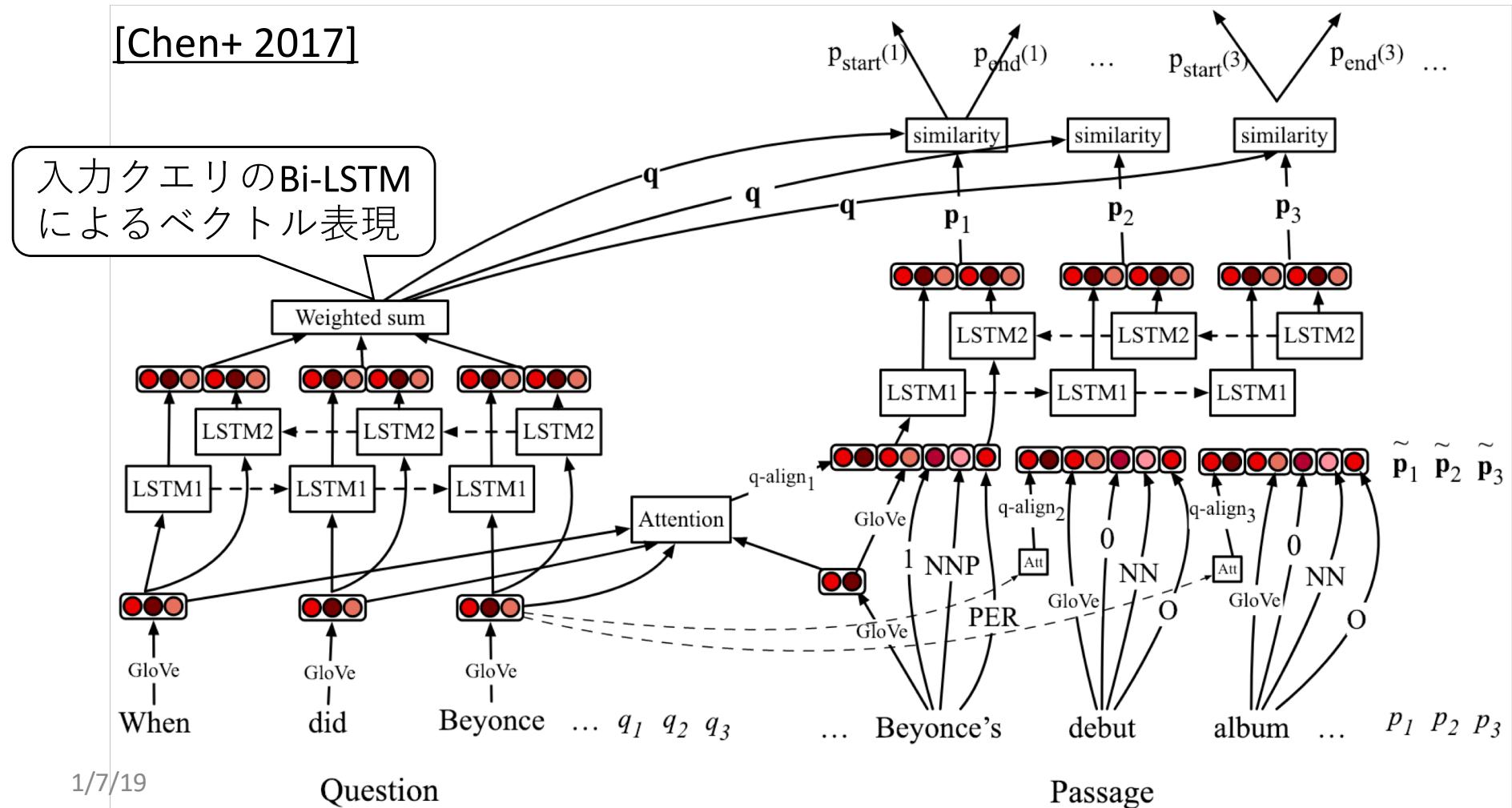
SQuAD datasets

[Rajpurkar+ 2018]

情報検索に基づくファクト型質問応答: ニューラル解答抽出 (2/2)

- 入力文書(パッセージ)の各単語について正解となるスパンの開始・終了単語となる確率を計算

[Chen+ 2017]



余談: 機械読解は人の読解能力を超えた?

- SQuAD データ・セットにおいて、人の回答スコアに匹敵する(あるいは上回る)システムが出現
<https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>

SQuAD1.1 Leaderboard

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) Google AI Language https://arxiv.org/abs/1810.04805	87.433	93.160
2 Sep 09, 2018	nlnet (ensemble) Microsoft Research Asia	85.356	91.202
3 Jul 11, 2018	QANet (ensemble) Google Brain & CMU	84.454	90.490

SQuAD2.0 Leaderboard

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Dec 13, 2018	BERT finetune baseline (ensemble) Anonymous	83.536	86.096
2 Dec 16, 2018	Lunet + Verifier + BERT (ensemble) Layer 6 AI NLP Team	83.469	86.043
3 Dec 15, 2018	Lunet + Verifier + BERT (single model) Layer 6 AI NLP Team	82.995	86.035

- Goldberg (2017)による指摘

解答不能な質問を含む

- Amazon Mechanical Turk で5問/2分で回答(時間制限あり)
- 人の誤りはスパンの範囲選択誤りが大半
- 質問の多くのはパターンマッチで解ける

質問応答タスクのデータセット

- **TREC QA データセット** [TREC, 1999-2004]
 - Text Retrieval Conference で人手で作成された質問と正解
- **WebQuestions** [Bernants+ 2013]
 - Web ユーザによる 5800 質問 (wh- で始まる factoid 型の質問) に対し, Freebase から人手で正解を抽出・付与
 - 意味解析結果 [Yin+ 2016] や複雑な質問を付与 [Talmor+ 2018]
- **TriviaQA** [Joshi+ 2017] <http://nlp.cs.washington.edu/triviaqa/>
 - 95K 質問-回答ペアに対して, 回答の知識源として平均6文書を収集したデータセット (650K 質問-回答-文書)
- **SQuAD** [Rajpurkar+ 2016], **WikiQA** [Yang+ 2015], **NarrativeQA** [Tomáš Kočiský+ 2018], **QuAC** [Choi+ 2018]

ファクトトイド型質問応答システムの評価

- Mean Reciprocal Rank (MRR) [TREC-8, 1999]
 - システムが返す解答候補リストと正解の一致度を計算

$$\text{MRR} = \frac{1}{N} \sum_{\substack{i=1 \\ \text{s.t. } rank_i \neq 0}}^N \frac{1}{rank_i}$$

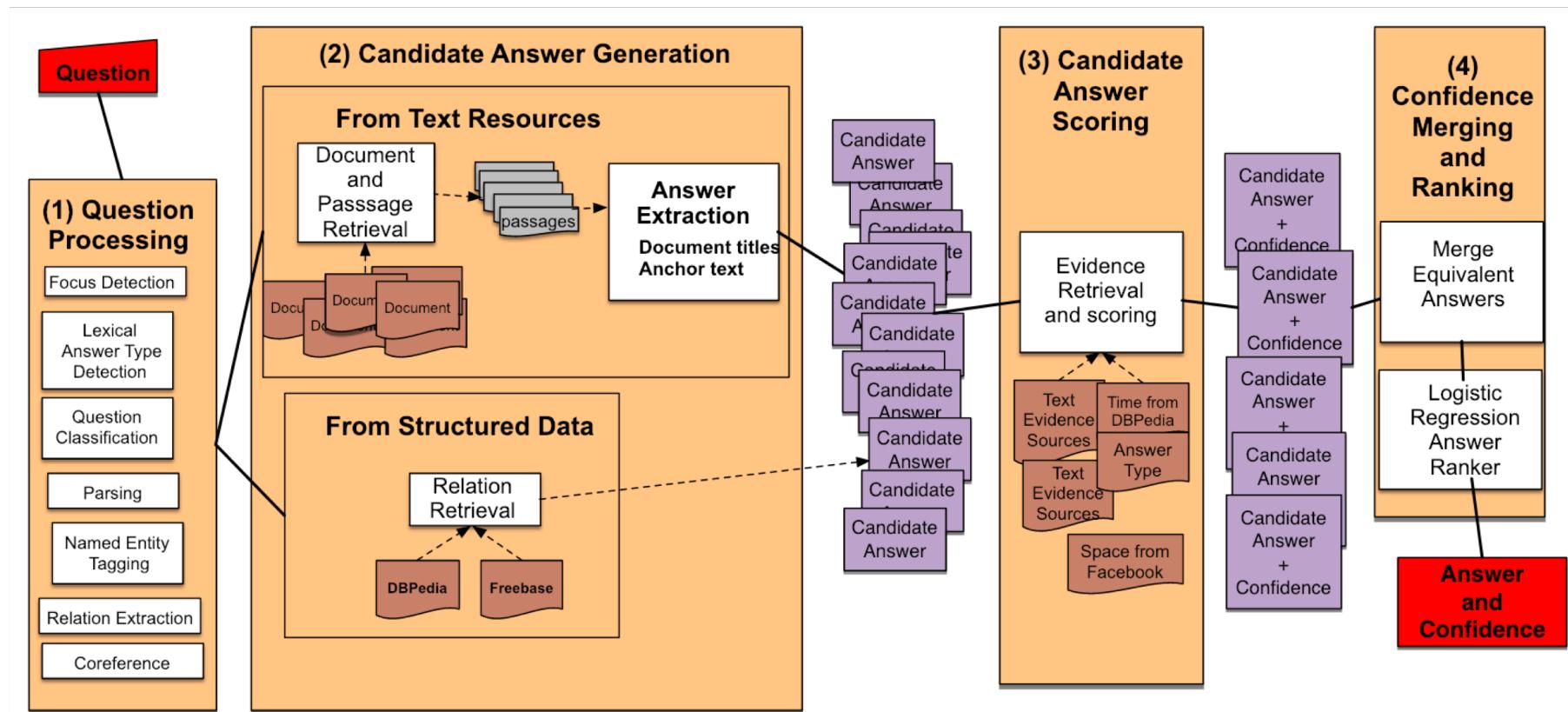
- 文書理解タスクの評価尺度 [Rajpurkar+ 2016]
 - Exact Match (EM)
 - F_1 : 正解の単語集合とシステムの出力の一致度 (BoW)

IBM Watson (1/2)

- IBM が米国の人気クイズ番組 *Jeopardy!* で人間に勝つために開発した質問応答システム (DeepQA)
 - 情報検索に基づく質問応答 + 知識に基づく質問応答
 - チェスで人間を越えることを目指した DeepBlue に続くグランドチャレンジとして2007年に開始 (2009年公表)
- 2011年2月に *Jeopardy!* で人間のチャンピオン2名と対戦し、勝利
 - 知識源として2億ページのテキスト ([Wikipedia](#)等) を利用
 - 音声認識は行わず、問題文はテキストで与えられる
 - ボタンを押したときの正答率は9割弱(チャンピオン並)

IBM Watson (2/2)

- IBM が米国の人気クイズ番組 *Jeopardy!* で人間に勝つために開発した質問応答システム (DeepQA)
 - 情報検索に基づく質問応答 + 知識に基づく質問応答



Wisdom X

<https://wisdom-nict.jp/>

- NICT が開発した約40億件のウェブ文書を知識源とする質問応答システム
 - ノンファクトトイド型の質問(なぜ型質問 [Oh+ 2012, Oh+ 2013], どうなる型質問[Hashimoto+ 2012, Hashimoto+ 2014]等)をサポート

例) どうなる型質問応答の連鎖による仮説形成

Q: 地球温暖化が進むとどうなる → 海水温が上昇する → 腸炎ビブリオ菌が増える → 食中毒が増える

The screenshot shows the Wisdom X search interface. The search bar contains the query "なぜ日本はデフレに陥ったのか?". Below the search bar, there is a link to "検索結果 [200件]" (Search results [200 items]). The results are listed in a card-based format:

1. A card from "kabu.com" discusses the reasons behind Japan's deflation, mentioning consumption tax, imports, and economic structure. It includes a link to <http://fathering.okwave.jp/qa8332142.html>.
2. A card from "blogos.com" discusses the causes of deflation, mentioning population aging and its impact on the economy. It includes a link to <http://blogos.com/article/36272/>.

対話システム

Dialogue System

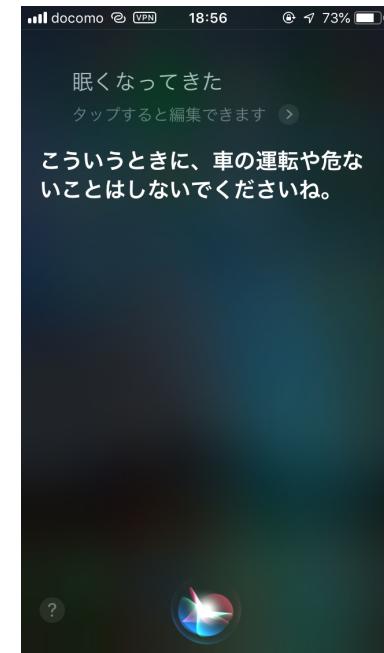
注) 本講義では自動音声認識 (ASR) は扱いません

対話システム

- タスク指向対話システム
(Task-oriented dialogue system)
 - 特定タスクを完了するのに必要な情報を提供
 - 目標: なるべく短いやり取りで情報を返す
- 雜談対話システム
(Chatbots)
 - 人に代わって人の話し相手となる(傾聴, 娯楽)
 - 目標: なるべく長くやり取りを続ける



質問応答システムと異なり
インタラクションがある



対話システム

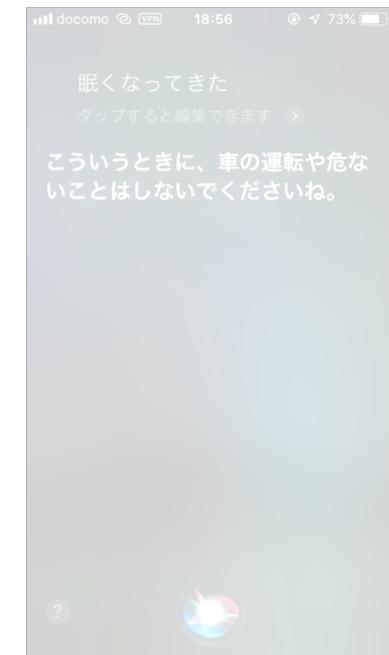
- タスク指向対話システム
(Task-oriented dialogue system)
 - 特定タスクを完了するのに必要な情報を提供
 - 目標: なるべく短いやり取りで情報を返す



1/7/19

- 雜談対話システム
(Chatbots)

- 人に代わって人の話し相手となる (傾聴, 娯楽)
- 目標: なるべく長くやり取りを続ける



27

フレームに基づくタスク指向対話システム

- タスクに必要な情報(フレーム)を収集するよう対話
 - GUS [Bobrow+ 1977]: フライト予約のための対話システム

Slot	Type
ORIGIN CITY	city
DESTINATION CITY	city
DEPARTURE TIME	time
DEPARTURE DATE	date
ARRIVAL TIME	time
ARRIVAL DATE	date

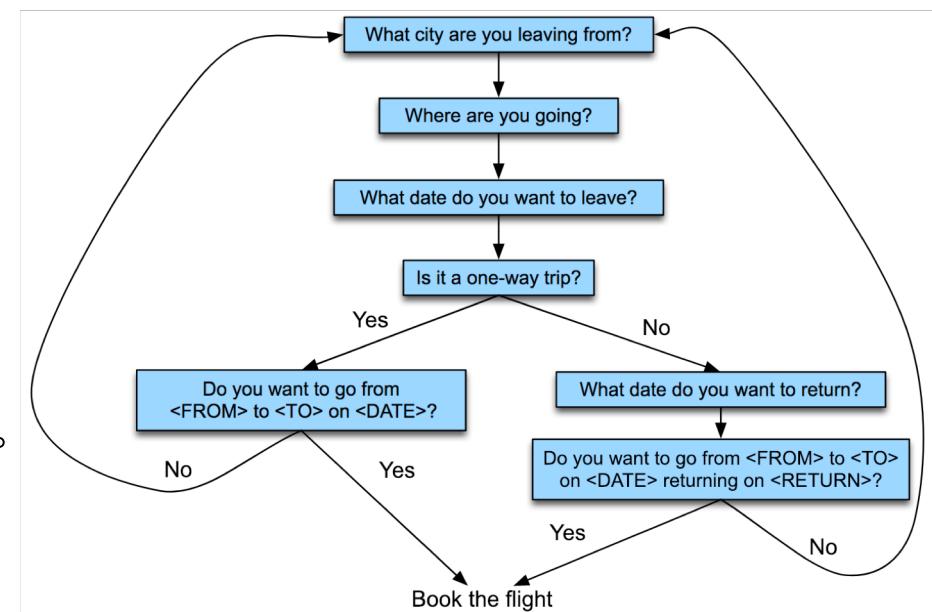
GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?
Client: I want to go to San Diego on May 28
GUS: What time do you want to leave?
Client: I must be in San Diego before 10 am
GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?
Client: What is 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: 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

- VoiceXML: タスク指向対話システムの実装枠組(略)

フレームに基づくタスク指向対話システム: 発話制御

- 有限状態オートマトンに基づく対話制御
 - フレーム中の各 slot の値を埋める質問を用意
 - ユーザはシステムの質問に順次応答 (対話システムの制御を行うための普遍的なコマンド (help, etc.) は入力可)

- 対話の主導権
 - システム主導:
会話の主導権をシステムが握る
 - 混合主導:
ユーザが複数 slot を埋める応答をした場合に、質問をスキップ
 - ユーザ手動: CUI など



フレームに基づくタスク指向対話システム: Slot filling に基づく自然言語理解 (1/4)

- 入力発話について、以下の3タスクを解く
 - ドメイン分類 (多クラス分類):
タスクに関するドメインから適切なドメインを選択
 - ユーザ意図推定:
ドメインでユーザが解決しようとするタスクを推定
 - Slot filling (スロットフィリング):
ユーザの応答からフレームの slot の値を抽出

例) *Show me morning flights from Boston to London on Tuesday*

ドメイン: AIR-TRAVEL

ユーザ意図: SHOW-FLIGHTS

ORIGIN-CITY: Boston

ORIGIN-DATE: Tuesday

ORIGIN-TIME: morning

DEST-CITY: London

フレームに基づくタスク指向対話システム: Slot filling に基づく自然言語理解 (2/4): 手法

- ルールベースに基づくドメイン・意図推定
 - パターン
 - wake me (up) | set (the|an) alarm | get me up → SET-ALARM intent
 - 形式文法 [Ward+ 1994]

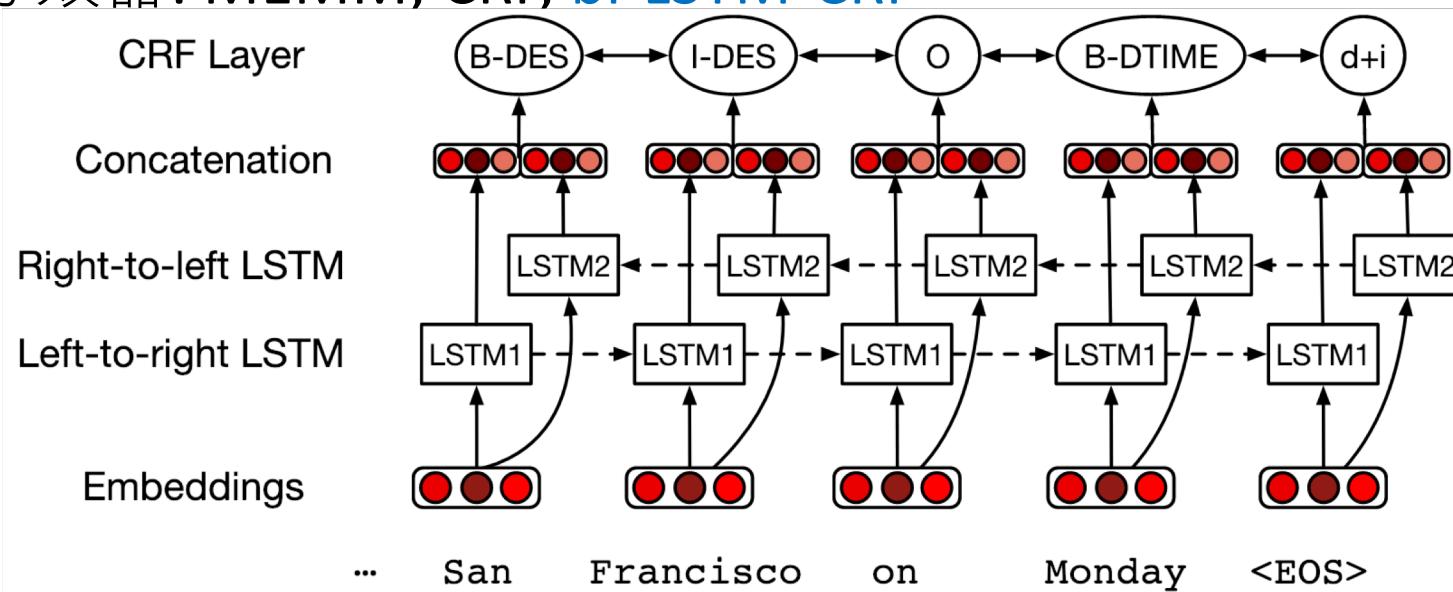
SHOW	→ show me i want can I see ...
DEPART_TIME_RANGE	→ (after around before) HOUR morning afternoon evening
HOUR	→ one two three four... twelve (AMPM)
FLIGHTS	→ (a) flight flights
AMPM	→ am pm
ORIGIN	→ from CITY
DESTINATION	→ to CITY
CITY	→ Boston San Francisco Denver Washington
- 教師あり学習に基づくドメイン・意図推定

フレームに基づくタスク指向対話システム: Slot filling 基づく自然言語理解 (3/4): 手法

- パイプラインによる教師あり slot filling
 - ユーザ発話によりどの slot が埋まるかを分類
 - slot のタイプに対応する分類器で系列から slot の値を抽出
- 系列ラベリングによる教師あり slot filling

I want to fly to San Francisco on Monday afternoon please
O O O O O B-DES I-DES O B-DEPTIME I-DEPTIME O

- 分類器: MEMM, CRF, bi-LSTM-CRF



フレームに基づくタスク指向対話システム: Slot filling に基づく自然言語理解 (4/4): 評価

- Slot Error Rate

$$\text{Slot Error Rate for a Sentence} = \frac{\# \text{ of inserted/deleted/substituted slots}}{\# \text{ of total reference slots for sentence}}$$

例) *Make an appointment with Chris at 10:30 in Gates 104*

<u>システムの出力</u>	Slot	Filler	Slot error rate = 1/3
	PERSON	Chris	
	TIME	11:30 a.m.	
	ROOM	Gates 104	

- Task Error Rate: 目的タスクの達成率

タスク指向型対話システムの評価

- ユーザ満足度 (user satisfaction rating) [Walker+ 2001]

TTS Performance	Was the system easy to understand ?
ASR Performance	Did the system understand what you said?
Task Ease	Was it easy to find the message/flight/train you wanted?
Interaction Pace	Was the pace of interaction with the system appropriate?
User Expertise	Did you know what you could say at each point?
System Response	How often was the system sluggish and slow to reply to you?
Expected Behavior	Did the system work the way you expected it to?
Future Use	Do you think you'd use the system in the future?

- (自動計算可能な)代替スコア
 - Task completion success: 正解 filler を得た slot の割合
 - Efficiency cost: 対話時間, システム発話数 [Polifroni+ 1992]
 - Quality cost: ASR の成功率, 応答速度, slot error rate

タスク指向対話システムをどう設計すれば良い？

User-centered principle [Gould+ 1985]

- 潜在ユーザとタスクの把握，既存システムの調査，人が同タスクをどう解くか分析
- プロトタイプの作成
- 実ユーザの検証を含むシステムの漸次開発

対話システム

- タスク指向対話システム
(Task-oriented dialogue system)
 - 特定タスクを完了するのに必要な情報を提供
 - 目標: なるべく短いやり取りで情報を返す



- 雜談対話システム
(Chatbots)

- 人に代わって人の話し相手となる (傾聴, 娯楽)
- 目標: なるべく長くやり取りを続ける



雑談は人の会話の半数を占めており、対話システムは両種類の対話をを行うことが求められる

- ユーザの発話か雑談意図を認識する研究も [Akasaki+ 2017]

雑談対話システム (Chatbots)

- 人との間の雑談を計算機で模倣することを目的とする対話システム
 - 対話システムにおけるユーザエンゲージメントの向上
 - 傾聴対話, カウンセリング [Weizenbaum 1966]
- ルールベース Chatbot:
ELIZA [Weizenbaum 1966], PARRY [Colby+ 1971]
- コーパスに基づく Chatbot:
与えられた発話(履歴)に対する応答生成問題を解く
 - 情報検索ベース [Isbell+ 2000, Jafarpour+ 2009, Leuski+ 2011, Yan+ 2016]
 - 機械翻訳ベース [Ritter+ 2011, Hasegawa+ 2013] (+ ニューラル系論文)

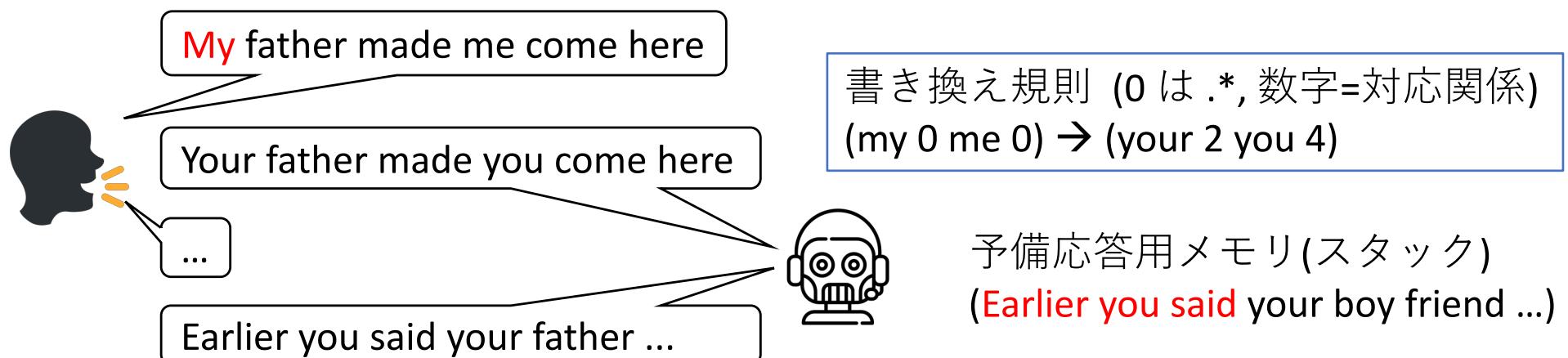
単純なルールに基づく Chatbots (1/2): ELIZA と PARRY

- ELIZA [Weizenbaum 1966]:
来談者中心療法 (Rogerian psychotherapy) における会話を模倣する Chatbot
 - 注) 会話をを行う上で世界知識が必要ない

```
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  
..
```

単純なルールに基づく Chatbots (2/2): ELIZA と PARRY

- ELIZA [Weizenbaum 1966]:
来談者中心療法 (Rogerian psychotherapy) における会話を模倣する Chatbot
 - 会話を行う上で世界知識が必要ない
- アルゴリズム: キーワードに基づくパターン応答
 - 発話から応答を決めるキーワードを認識しパターン選択



1/7/19 PARRY は話者の感情をモデル化し Turing test を突破

コーパスに基づく Chatbot: 情報検索に基づく手法

- 大規模会話ログ(発話-応答ペア) C を知識源として,
入力発話 q に近い発話 t に対する応答を返す
 - Twitter や Weibo などの大規模ユーザ発話ログを利用
 - 類似度は発話ベクトル(tf-idf や分散表現)間の内積で計算

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

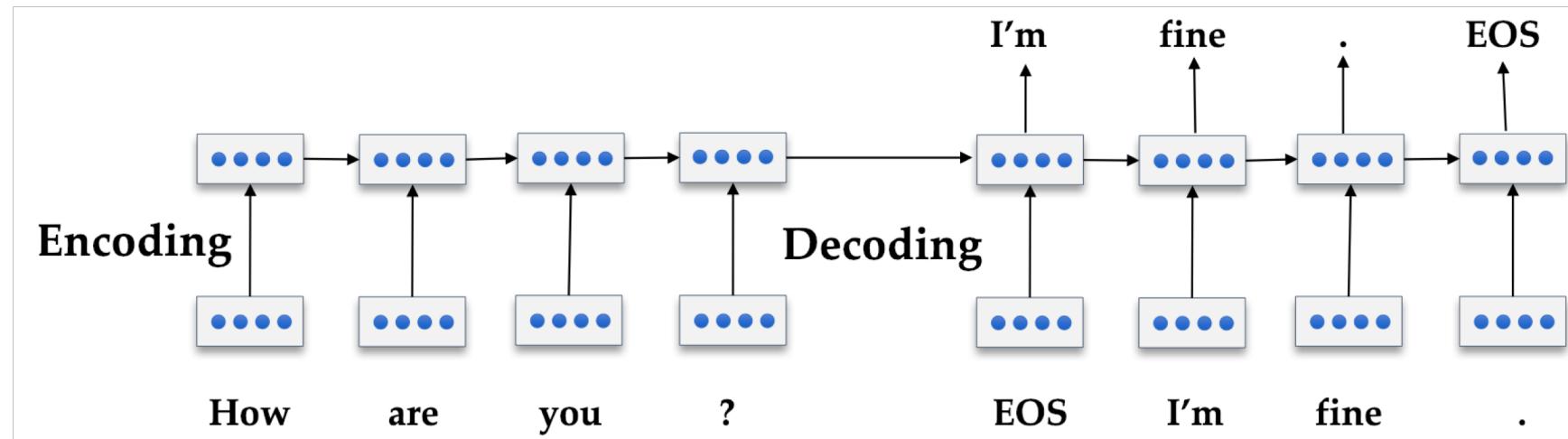
- あるいは入力発話に近い発話 자체を返す

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

言語としてまともな応答が可能だが話が噛み合わない場合も

コーパスに基づく Chatbot: 機械翻訳に基づく手法 (1/2)

- 入力発話を原言語、応答を目的言語とみなして、
機械翻訳モデルで応答生成
 - 統計的機械翻訳 [Ritter+ 2011, Hasegawa+ 2013]
 - ニューラル機械翻訳 [Shang+ 2015, Vinyals+ 2015, Sordoni+ 2015]



コーパスに基づく Chatbot: 機械翻訳に基づく手法 (2/2)

- 入力発話を原言語、応答を目的言語とみなして、
機械翻訳モデルで応答生成
 - 統計的機械翻訳 [Ritter+ 2011, Hasegawa+ 2013]
 - ニューラル機械翻訳 [Shang+ 2015, Vinyals+ 2015, Sordoni+ 2015]
- 課題:
 - 典型応答への対処 [Li+ 2016]
 - 過去の会話履歴の考慮 [Lowe+ 2017]
 - 発話間の一貫性 [Li+ 2016, Li+ 2017]
 - 話者の感情や発話状況のモデル化 [Hasegawa+ 2013, Sato+ 2017]

生成でまともな応答を返すことは現状かなり厳しい

Chatbot の評価

- 基本的には人手評価
 - 雑談では発話に対する応答は一意に定まらないため
- 自動評価への試み
 - ADEM [Lowe+ 2017]:
発話-応答ペアに対し人手でつけた応答の適切度を元に,
評価モデルを自動学習
 - Adversarial evaluation [Bowman+ 2016]:
人の応答とシステムの応答を識別する分類器を学習し,
より人の応答と判断される応答が適当であるとみなす

対話システムにおける倫理的問題

[Henderson+ 2017]

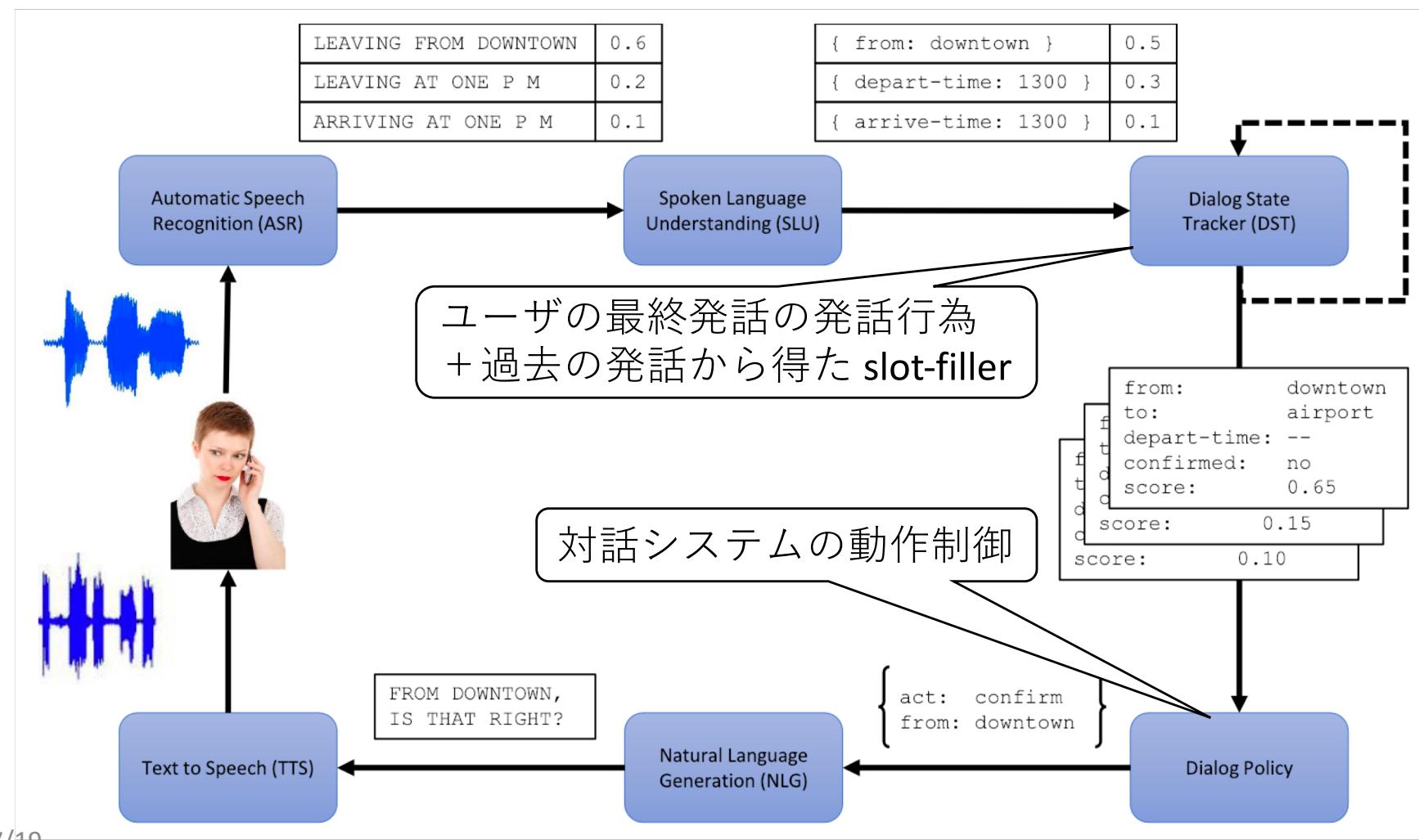
- 実ユーザの発話を扱う対話システムではバイアスやプライバシーに絡み様々なる倫理的問題が生じる
 - Twitterなど既存の対話ログから学習した Chatbot は、元データに含まれるバイアスを保持(あるいは強調)
 - ユーザの発話から学習を行うリスク(攻撃)
cf. Microsoft Tay: <http://time.com/4270684/microsoft-tay-chatbot-racism/>
Twitter ユーザによる「調教」でヘイトスピーチを学習
 - ユーザの発話を記録するリスク
cf. password など秘密情報
- 現状の知的対話エージェント多くは女性名で運用されている(家事従事者としての女性を想起)

より高度な対話システムに
に向けて

Advanced Dialogue System

対話状態を考慮した対話システム

- [Williams+ 2016]



対話行為 (Dialogue Acts) (1/5)

- 対話行為 = 発話行為 [Austin 1962] + 基盤化 (grounding)
 - 発話行為 (speech act): 話者が発話により意図する行為

[Bach+ 1979]

Constatives	事実の陳述 (answering, claiming, confirming...)
Directives	行為の指示 (advising, asking, inviting ...)
Commissives	行為の約束 (promising, planning, vowing, ...)
Acknowledgements	話者の態度 (apologizing, thanking, ...)

- 基盤化: 発話に含まれる情報を正しく理解し, 共通基盤 (common ground; 話者が共有する情報)に追加

[Clark+ 1989]
による基盤化
の強さの分類

Continued attention	注視を続ける
Next contribution	話者の発話への応答 (疑問に対する回答等)
Acknowledgement	頷く, 相槌を打つ (continuer or backchannel)
Demonstration	話者の発話を言い換えたり補足する
Display	話者の発話をそのまま繰り返す

対話行為 (Dialogue Acts) (2/5): 基盤化の例

C₁: ...I need to travel in May.

A₁: And, what day **in May** did you want to travel? Display

C₂: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A₂: And you're flying into what city?

C₃: Seattle. Next contribution

A₃: And what time would you like to leave Pittsburgh?

C₄: Uh hmm I don't think there's many options for non-stop.

A₄: Right. There's three non-stops today.

C₅: What are they?

A₅: 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.

C₆: OK I'll take the 5ish flight on the night before on the 11th.

A₆: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

C₇: OK.

対話行為 (Dialogue Acts) (2/5): 例

- 会議のスケジューリングタスクにおける発話行為

Tag	Example
THANK	<i>Thanks</i>
GREET	<i>Hello Dan</i>
INTRODUCE	<i>It's me again</i>
BYE	<i>Alright bye</i>
REQUEST-COMMENT	<i>How does that look?</i>
SUGGEST	<i>from thirteenth through seventeenth June</i>
REJECT	<i>No Friday I'm booked all day</i>
ACCEPT	<i>Saturday sounds fine</i>
REQUEST-SUGGEST	<i>What is a good day of the week for you?</i>
INIT	<i>I wanted to make an appointment with you</i>
GIVE_REASON	<i>Because I have meetings all afternoon</i>
FEEDBACK	<i>Okay</i>
DELIBERATE	<i>Let me check my calendar here</i>
CONFIRM	<i>Okay, that would be wonderful</i>
CLARIFY	<i>Okay, do you mean Tuesday the 23rd?</i>
DIGRESS	<i>[we could meet for lunch] and eat lots of ice cream</i>
MOTIVATE	<i>We should go to visit our subsidiary in Munich</i>
GARBAGE	<i>Oops, I-</i>

対話行為 (Dialogue Acts) (3/5): 例

- HIS レストラン推薦システムにおける対話行為

Tag	Sys	User	Description
HELLO($a = x, b = y, \dots$)	✓	✓	Open a dialog and give info $a = x, b = y, \dots$
INFORM($a = x, b = y, \dots$)	✓	✓	Give info $a = x, b = y, \dots$
REQUEST($a, b = x, \dots$)	✓	✓	Request value for a given $b = x, \dots$
REQALTS($a = x, \dots$)	✗	✓	Request alternative with $a = x, \dots$
CONFIRM($a = x, b = y, \dots$)	✓	✓	Explicitly confirm $a = x, b = y, \dots$
CONFREQ($a = x, \dots, d$)	✓	✗	Implicitly confirm $a = x, \dots$ and request value of d
SELECT($a = x, a = y$)	✓	✗	Implicitly confirm $a = x, \dots$ and request value of d
AFFIRM($a = x, b = y, \dots$)	✓	✓	Affirm and give further info $a = x, b = y, \dots$
NEGATE($a = x$)	✗	✓	Negate and give corrected value $a = x$
DENY($a = x$)	✗	✓	Deny that $a = x$
BYE()	✓	✓	Close a dialog

対話行為 (Dialogue Acts) (5/5): 隣接ペアに基づく対話制御

- 隣接ペア [Schegloff 1968]: 隣接する発話の対話行為の対
 - 会話分析 (conversational analysis)[Sacks+ 1974]で定義
Questions - Answering, Proposal - Acceptance (or Rejection), Complements - Downplayers, Greeting - Greeting など
- 例外:
side sequence [Jefferson 1972, Schegloff 1972] と **pre-sequence**

User: Can you make train reservation?

System: Yes, I can.

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

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

User: Going to Hong Kong.

subdialog (副対話)

System: OK, here are some flights

対話状態 (Dialogue States)

- Dialogue state tracker により 最終発話の対話行為と 対話履歴から得られる情報を対話状態として保持

User: I'm looking for a cheaper restaurant
inform (price=cheap)

System: Sure. What kind - and where?

User: Thai food, somewhere downtown
inform(price=cheap, food=Thai, area=centre)

System: The House serves cheap Thai food

User: Where is it?
inform(price=cheap, food=Thai, area=centre); request(address)

System: The House is at 106 Regent Street

- 対話行為の解釈は語用論的な要素を含む

Can you give me a list of the flights from Atlanta to Boston?

対話行為の認識

- 発話の対話行為の推定と slot filling は対話行為が slot の値に制約を与えることがあるため順に解く

I'd like Cantonese food need the Mission District
→ inform(food=cantonese,area=mission))

- 対話行為の推定: 分類問題(教師あり学習)
- slot filling: 系列分類 (p.31)
- 音声認識の誤り訂正の対話行為の推定は困難
 - hyperarticulation (過剰に明瞭な発音)

Dialog Policy (1/2):

確認と否定に基づく対話制御

- 基盤化において誤解が混入することを防ぐために
確認要求を行う

System: Which city do you want to leave from?

User: Baltimore.

System: Do you want to leave **Baltimore**?

明示的確認

User: Yes, and I want to go to Berlin.

暗黙的確認

System: When do you want to travel to **Berlin**?

- 発話が理解できていないことをユーザに示すために**棄却**したり(段階的に)具体化して問い合わせる

User: I need to be in New York in time for the first World Series game.

System: Sorry, I didn't get that. Please say the date you'd like to leave.

User: I wanna go on October fifteenth.

Dialog Policy (2/2):

対話状態に基づく対話制御

- 過去の対話行為の履歴に基づき、次の対話行為を決定

$$\hat{A}_i = \operatorname{argmax}_{A_i \in A} P(A_i | A_1, U_1, \dots, A_{i-1}, U_{i-1})$$

- 直前のやり取り以外を対話状態中のフレーム (slot-value) で近似

$$\hat{A}_i = \operatorname{argmax}_{A_i \in A} P(A_i | \text{Frame}_{i-1}, A_{i-1}, U_{i-1})$$

- パラメタは機械学習により推定

対話状態に基づく対話システムにおける 言語生成

- Content planning (何を発話するか = 意味表現)
+ sentence realization (どう発話するか)
 - 非語彙化された文の生成 (生成問題の単純化):
大規模コーパスから対話行為ごとに生成モデルを学習

What time on [depart_date] would you like to leave [depart_airport]?
 - 語彙化
content planning の結果得られた slot 情報で語彙化

What time on Oct. 5th 2000 would you like to leave Boston?
- 深層学習に基づく強力な生成モデルではフレーム
から直接、発話を生成することも可能 [Wen+ 2015]

まとめ

- 質問応答システム
 - Factoid / non-factoid 型質問
 - 知識に基づく質問応答と情報検索に基づく質問応答
 - 言語理解タスクのベンチマークとしての質問応答
 - 実システム: IBM Watson, NICT WISDOM X
- 対話システム
 - フレーム (slot-filling) に基づくタスク指向対話システム
 - 対話状態に基づくタスク指向対話システム
 - コーパスに基づく雑談対話システム (Chatbot)
 - 実システム: Siri, しゃべってコンシェル, Cortana, りんな...