

シンボル・グラウンディング あるいは記号の創発に向けて

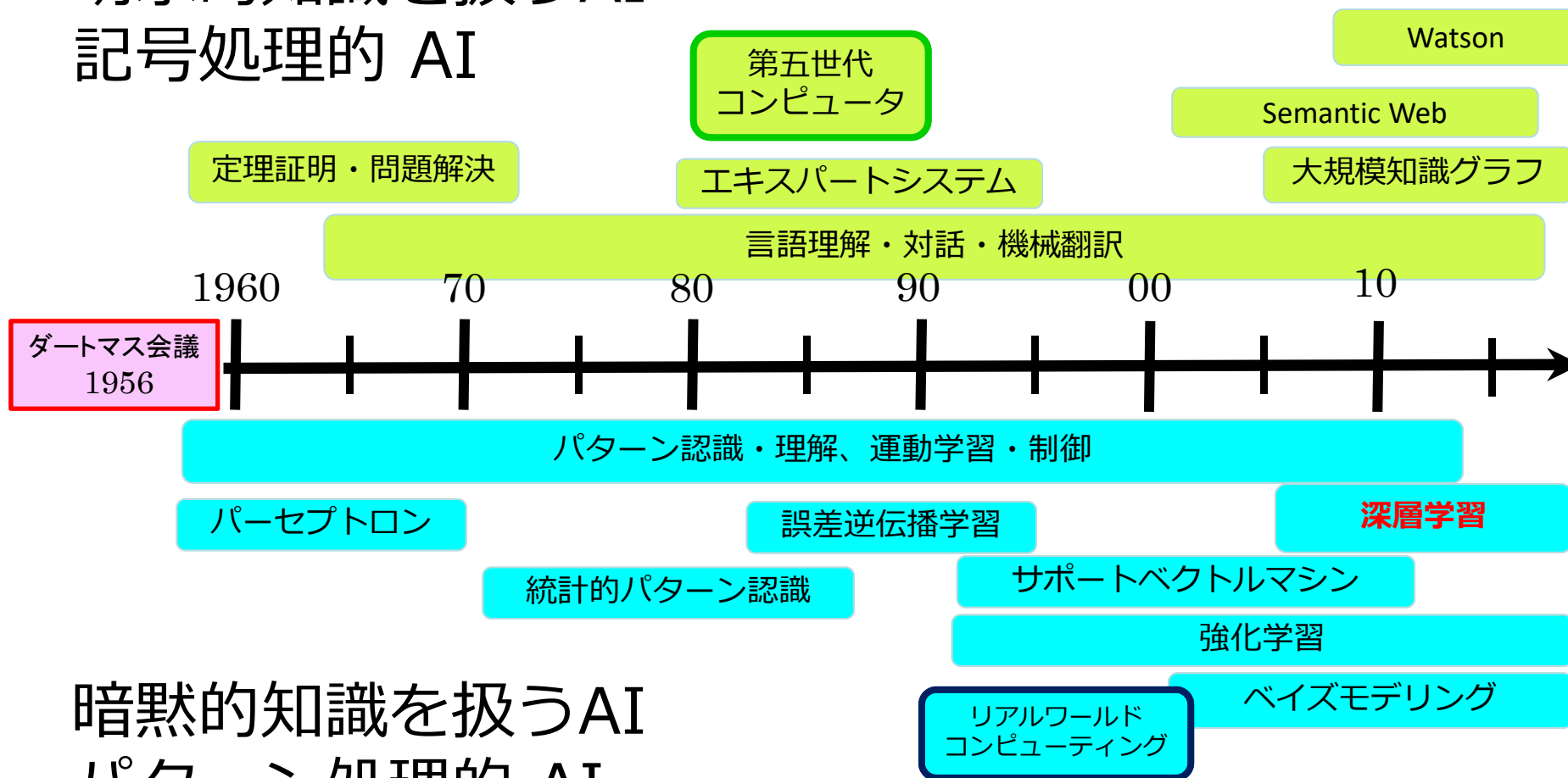
「深層学習の先にあるもの
～記号推論との融合を目指して」
公開シンポジウム
2018年1月22日

国立研究開発法人 産業技術総合研究所
人工知能研究センター
麻生 英樹

人工知能研究の二つの流れ

明示的知識を扱うAI
記号处理的 AI

それぞれの夏と冬



暗黙的知識を扱うAI
パターン处理的 AI
(ずっとAI とは呼ばれていなかった)

二種類の情報処理

意識に登る
外部記憶なども使った
記号処理

無意識の
パターン処理

長尾真「人工知能と人間」(1992)

- 「(画像や音声といった) アナログ情報がシンボル化以前のある種のイメージとして記憶されることは間違いないし、これが人間の知的機能であるシンボルと深層にある情動との間をつなぐ重要なキーとして働いていることも確かである。(中略)
ニューラルネットワークのような非シンボリック的取扱いを(記号処理的な人工知能に)組み込んだ総合的なシステムを検討することは重要なことである。しかしそれを行ったからといって、どこまで人間的なことを実現できるかはわからない。できる、できないと議論するのではなく、やってみる以外に方法はない。」(長尾真「人工知能と人間」(岩波新書), 1992)

安西祐一郎「心と脳」(2011)

- 「二つのモデル（コネクショニストモデルと記号表現モデル）の弱点を乗り越えた情報処理モデルは、まだ現れているとはいえない。この点を乗り越え、**記号を使う心のはたらきとそうでないはたらきが意識のうえと意識下でどう相互作用しているのか、とくに、人が知覚や運動の機能をはたらかせながら、その一方で記号によってことばを操れるのはなぜかという問題に答を与えることは、多くの認知科学者が挑戦してきた主要な課題の一つである。**」（安西祐一郎「心と脳－認知科学入門」（岩波新書），2011）

柔らかな記号に向けて (1988)

- 硬い記号
 - 現在の計算機が操作する記号
 - 組み合わせ構造の生成や操作が得意
- 柔らかな記号
 - 人間の脳の中で
分散表現された記号
 - 連想と組み合わせ構造生成の両方ができる



[麻生 1988]

脳と記号処理

- 脳は連想ベースのパターン情報処理装置
- 記号処理は本来は苦手
 - 記憶、計算、先読み計画、などは
計算機のほうがはるかに得意
- だから紙などの外部記憶装置や
計算装置を発明した

- それなのに記号的推論や言語を使っている
- それが人間の強みになっている

パターン処理と記号処理の統合

- 脳の情報処理原理の観点から [JNNS 2017プレイベント]
 - ボトムアップとトップダウン
 - 識別モデルと生成モデル
 - ベイズ的順 – 逆計算
 - 予測符号化、自由エネルギー最小化
 - Auto-encoding Variational Bayes
 - Bayesian Deep Learning
- もう少し工学的（人工知能的？）な観点から
 - 記号的なシミュレーションの、
プランニングの利用（計算削減）
 - 機械学習における事前知識・領域知識の活用
 - 人間と相互理解できる人工知能、XAI
 - 言語記号の使用

脳と記号処理

- 知能の目的は生存 = 報酬獲得 と仮定する
- 知能のコアはモデルなし強化学習
- 報酬予測・獲得のために
状況の認識や分類（教師なし学習）や、
精密な制御（教師あり学習）が発達した
→ モデルベースの強化学習の併用
- 記号処理・操作もそのために発達した
→ シミュレーションの高速化、定性推論？
- その能力が言語にも転用された？

- DeepMind は深層強化学習にフォーカス

言語記号の使用に向けて

いろいろなタスク設定

- 大規模物体検出・認識
Redmon, J. and Farhadi, A.: YOLO9000: better, faster, stronger, arXiv:1612:08242, 2016
- 画像のキャプション
Vinyals+: Show and tell: a neural image caption generator, 2014
Xu+: Show, Attend and Tell: Neural image caption generation with visual attention, 2015
- 画像生成
Mansimov+ : Generating images from captions with attention, 2016
- Visual Q&A
(<http://http://www.visualqa.org/challenge.html>)
Goyal+: Making the V in VQA matter: elevating the role of image understanding in visual question answering, 2017

※参照文献は多くの研究事例の中のひとつ

言語記号の使用に向けて

いろいろなタスクの設定

- ビデオのラベリング、キャプション
Natsuda+: Generating Video Description using Sequence-to-sequence Model with Temporal Attention, COLING 2016
- Video Story Q&A
Kim+: DeepStory: video story QA by deep embedded memory networks, 2017
- ビデオ生成
- テキスト \leftrightarrow 動作生成
- Semantic Parsing
テキスト \leftrightarrow 形式意味表現、データベース検索式
- 試験問題への回答
- スポーツの試合や将棋の解説生成
- 時系列データの説明
- . . .

実世界 Q&A タスク

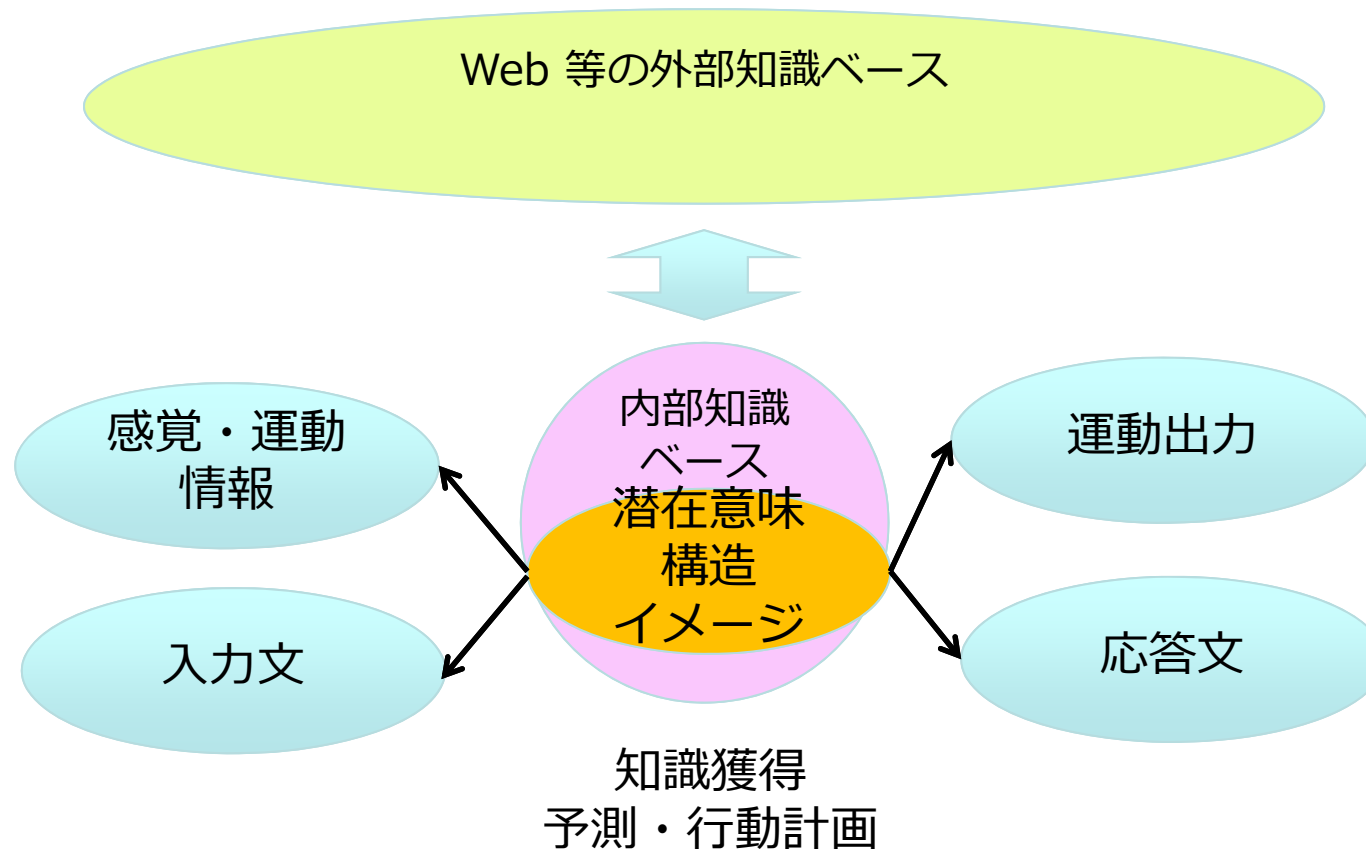
- 感覚・運動情報に関する質問応答
- 質問応答を通じた知識獲得
- 知識を使って適切に行動

- 強化学習 (Theory of Fun) や模倣学習との組み合わせ？



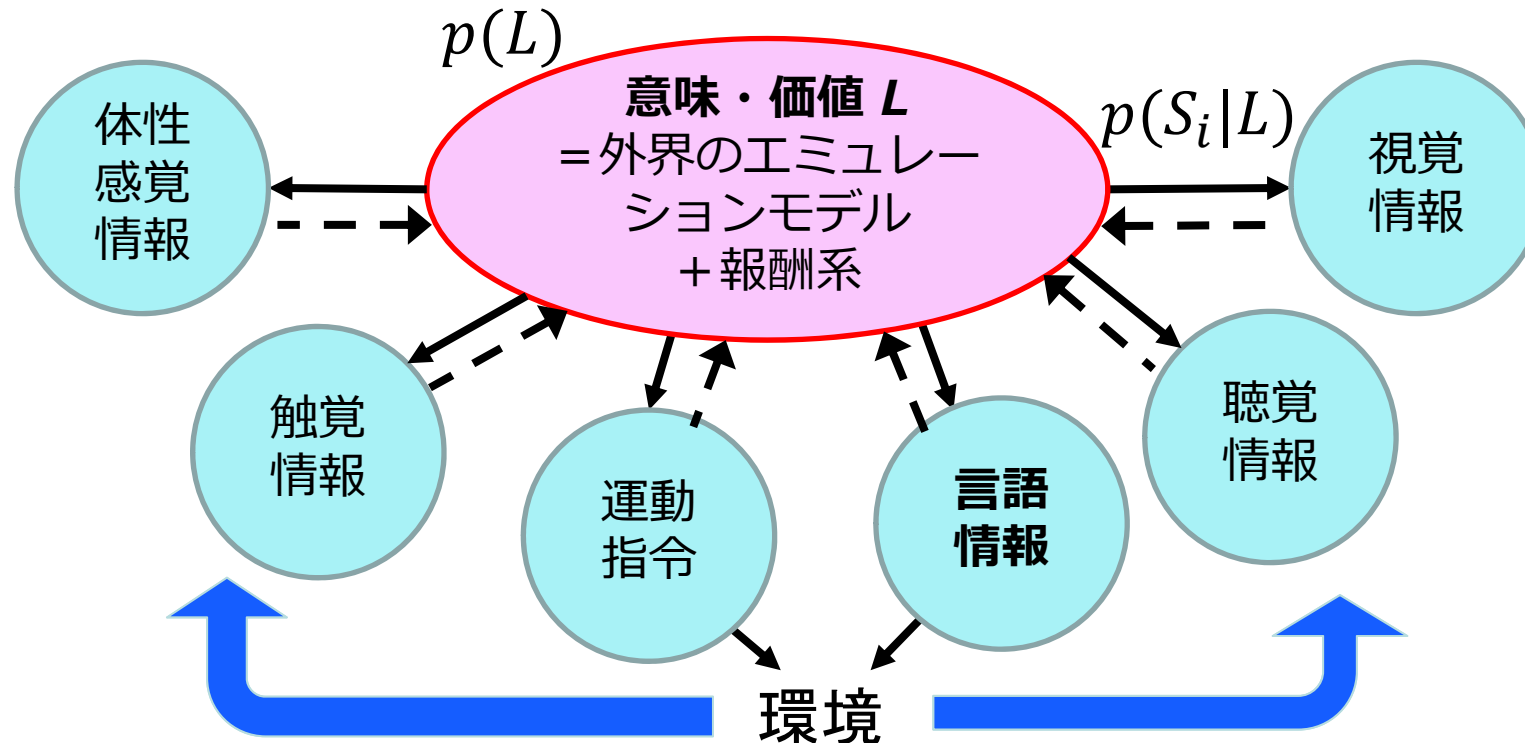
対話から学習する事情通ロボット
[Asoh+ 1997]

実世界 Q&A タスク



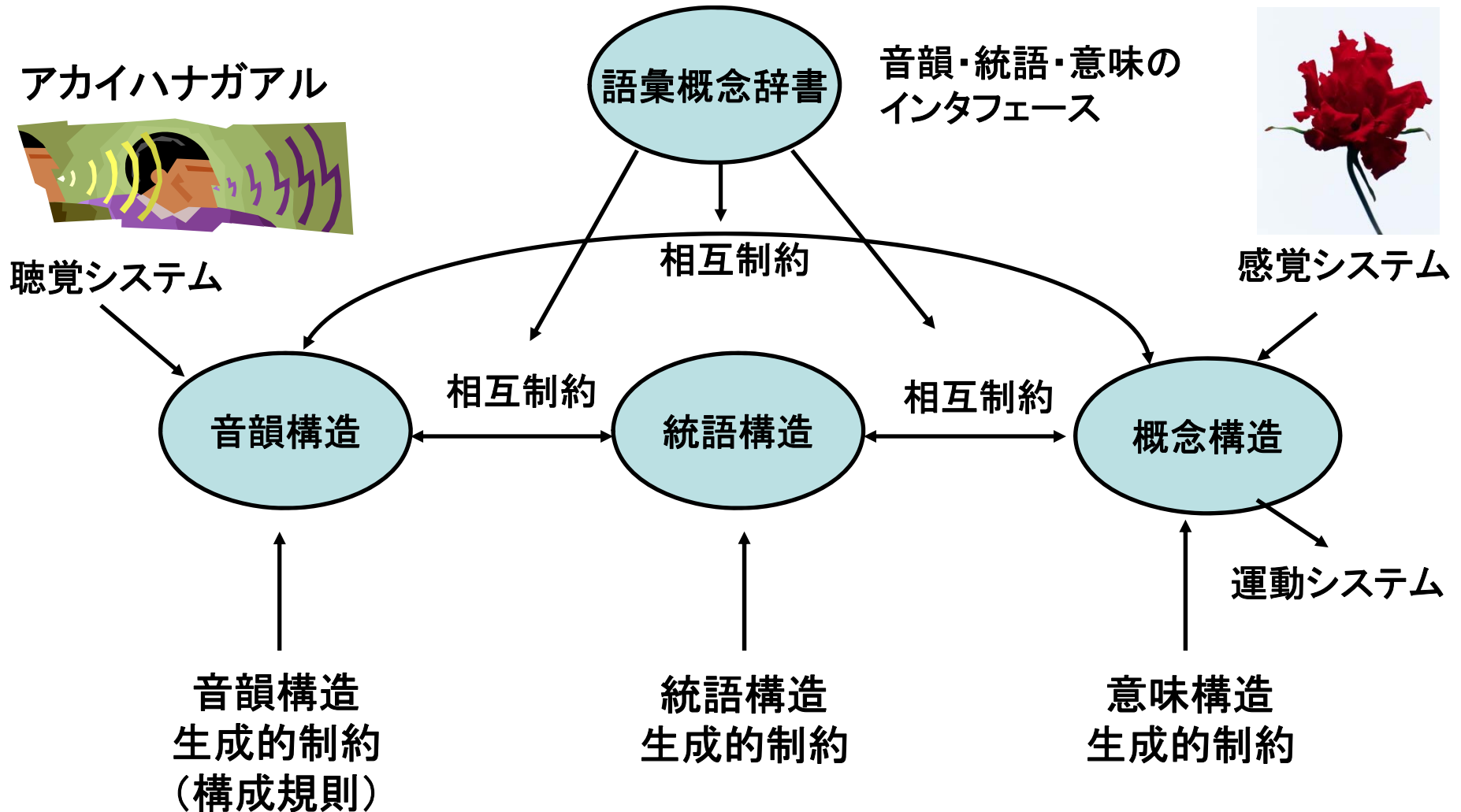
ベイズ的な言語学 [2012 麻生]

- マルチモーダルな生成モデルによって
言語、視覚、聴覚、触覚、を統合する
- モダリティをつなぐ潜在変数 = 意味
- S_i を個別にモデリングするよりも簡単になる可能性



言語に関する心的構造

[Jackendoff 2005 など]



何がどこまでできているのか？

- 名詞的概念、形容詞的概念、名詞句的概念 (entity) のグラウンディング
- 集合概念のグラウンディング
- 動詞的概念、副詞的概念、動詞句的概念 (event) のグラウンディング
- 動作系列 (event が並んだ現象) のグラウンディング
- より長い記憶・文脈を含めたグラウンディング
- 体系的なタスク設定と統合的な解法は？
- 人間の発達もあまりはっきりしていない？

集合概念の扱い

- VQA での How many 質問の正答率は他に較べてまだ低い
- 「2人の画像」、「3人の画像」、・・・から「5人の画像」を作れるか？
- 「1人だけ帽子を被っていない3人の画像」を作れるか？

時間や空間の扱い

- イベントの系列
 - 例：「出張」概念
出発－移動－到着－宿泊－仕事－食事－・・・
 - イベントの組み合わせは定長で可変
- 記述の粒度
 - 概念形成のノンパラメトリック的な自由さ

構造的な意味表現は必要か？

- End-to-End (+メタ学習) で良いのか？
- 構造的な意味表現は必要か？
- 事前知識として何が入られるのか？
- Semantic Parsing
 - テキスト \Leftrightarrow 意味表現 (λ式、DCSなど)
- Visual Q&A と Semantic Parsing の融合？
- 予測・プランニング・シミュレーション
強化学習との融合？

データが自然に集まる仕組み

- マルチモーダルな対話データが集まるサービスを通じたデータの収集
 - スマートスピーカーには耳しかない
 - ロボットのサービスは？
- 動画のテキスト検索？
- 実世界 Q&A？
- 地理空間情報の変化検出と解釈？

ある日の対話

H: 昨日上野でAさんを見かけたよ

R: そうですか

H: Bさんと一緒に食事していた

R: そうですか

H: AさんとBさんは仲がいいのかな？

R: いえ、そうではないみたいです

H: どうして？

R: わたしの得た情報によると・・・

- 人間には普通にできるのに、現在の対話システムにはまだまだ遠いレベルのように見える
- そこに至る発達段階もよく見えてはいない
- 模倣学習 + 強化学習から概念（教師なし学習）、言語へ？