

# 言語解析論

講師 竹内孔一

# 内容

- 意味解析と語義曖昧性解消

# 意味に関係しそうな研究分野

## 哲学

アリストテレス  
カント  
ヴィトゲンシュ  
タイン  
ヒューム  
意味とは?  
知識とは?

## 論理学

フレーゲ  
ゲーデル  
タルスキー  
意味を  
形式的に  
扱おう!  
推論しよう!

## 計算機科学

Turing  
Church, Rosser  
計算すると  
は何ですか?  
計算の世界は  
1つしか無い

## 人工知能

Shank  
Winograd  
高木  
言語知識を  
実体世界解釈に  
近づけて解く!

## 理論

## 手法

ベイジアン  
ネットワーク

Pearl  
Humeの因果を  
取り扱う手法

## 形式意味論 数理論理学

ヒルベルト  
ラッセル  
矛盾無く  
公理から  
導ける?

## 言語学・意味論 レキシコン

Levin,  
Jackendoff  
Pustejovsky  
影山, 池上  
西山  
言語表現を手  
掛かりに意味を  
考えよう

## オントロジー

ヨーロッパでの  
数々の構築例  
意味を形式的に  
書いて使おう  
(分野依存)

## 設計学

部品を機能  
で記述して  
シミュレーション  
自己修復  
コピー機販売

## 統計的手法 Collobert他

Deep Neural  
Network  
Dirichlet  
Distribution  
意味も潜在  
変数として  
扱おう

などなど

# 言葉の意味とは何か？

- 言葉の意味

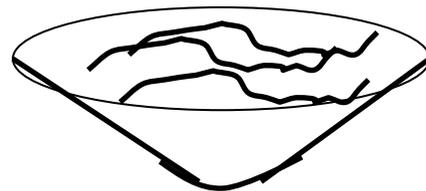
ソシュールの考え方

– *signifiant* (シニフィアン) と *signifié* (シニフィエ)

– その言葉(*signifiant*) と言葉が参照する先(*signifié*)

- シニフィアン 「うどん」「Udon needle」 ある内容を指す言葉
- シニフィエ 「うどん」という概念内容そのもの

「うどん」



「Udon needle」

「饅頭」

# 言葉の意味を扱う(1/2)

- 方法2つ
    - (1) 参照先の概念世界の制約を書く
    - (2) 言葉の規則の組を処理
  - (1) 参照先の概念世界の処理
    - 知識工学, オントロジーの分野
      - 言葉の外側の世界
    - 例)
      - 「A君は10kmの道を30分で走りました. 時速はいくら?」
      - 「東京まで片道で切符を買いたいのですが?」「何時の新幹線ですか?」
- 知識工学の講義で扱います(この講義ではしない)

# 言葉の意味を扱う(2/2)

- (2) 言葉の規則の組を処理

ここでの扱う

- 言語解析, 意味役割

- 単語間の関係の整理

- 例) Winograd(1972), Allen (1987)

- 「三角形」が (1) の概念として定義されてるとき

「赤い三角形」: S="三角形", Color(S,赤い) または S.color (RED)

「彼が昨日拾った三角形」: 拾う(彼, 三角形) & Time(昨日)

形式的表現

Take(Agent: 彼, Theme: 三角形) & Time(昨日)

Take(e) & Agent(e, 彼) & Theme(e, 三角形) & Time(e, 昨日)

これが意味役割 (つまり述語に対する項の種類)

# 言葉の意味処理の手法

- 意味処理の手順(古典)
  - 文書の構造(構文解析)から, 単語間の関係を整理((2)の処理)をして, システム内で必要とする概念(既に決めた概念)にマップさせて(1) 人の言葉がシステム内の概念の何に当たるのかを探す
- 近年
  - (2)の構造化も含めて統計的学習により(1)のシステムで必要とする概念に直接マッチさせる
  - 参照: 例) Berant, J. and Liang, P., Semantic Parsing via Paraphrasing Proceedings of ACL2014.

質問文      How many ships are designed by Nat Herreshoff?

論理的な検索式            count(BoatDesigner.NatHerreshoff)

DB依存の式      [Wikipedia DB](#)  
[FreeBase](#)

# 練習

- 「彼が落とした本」を形式的な表現(述語論理に近い形)で書いてみよう

# 文の意味処理(ここでは)

- 手順
  - 構文解析または係り受け解析
  - 単語間の意味的關係を同定
    - 意味役割解析
- 文の構造は何か?
  - 分析してみよう
- 概念フレーム曖昧生解消

# 分析

自然言語処理

言語学



知識処理(人工知能)

非文かどうかを手かがり  
として語・句・文・発話の意味

目的に合わせてある概念を  
形式化→部品化

パソコンのプリンタは今日到着するだろう

名詞句  
解析

「AのB」  
「AはBだ」西山  
名詞項構造 (影山)  
NomBank

命題

述語の概念化

LCS (影山)  
GL(Pustejovsky)  
FrameNet,  
VerbNet  
動詞項構造  
シソーラス

判断(モダリティ)

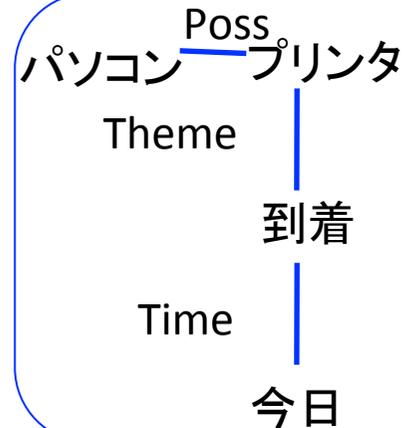
時間

時相論理

松吉他  
モダリティ  
コーパス

フレーム (Minsky)  
意味ネットワーク

(パソコン  
Poss: プリンタ)



状態遷移

FSBモデル

部品の  
挙動を  
状態遷移  
で記述

故障診断  
代替機能

1980~96

# 名詞句の分析

- 理論&人工知能
  - 西山の名詞句分析
  - 高木の自然言語理解
  - 影山のGLベース

## 高木(1988)

「あの車は赤い」「赤色の車」「色が赤い車」

車 ( の )色 ( 赤い )  
 ○ = ● => ◎ <- \* = ○ = ● => ◎ <- ○ = ● => ◎ <- [赤]  
 CAR の COLOR POSS HUE EQ  
 POSS

### 単に属性-属性値のオブジェクト

(Name: 車  
属性: 色: 赤)

「色」「値段」「高さ」  
 「重さ」は属性として  
 別扱い

## 西山(2003, 2011)

指示的名詞  
 非指示的名詞 (属性)

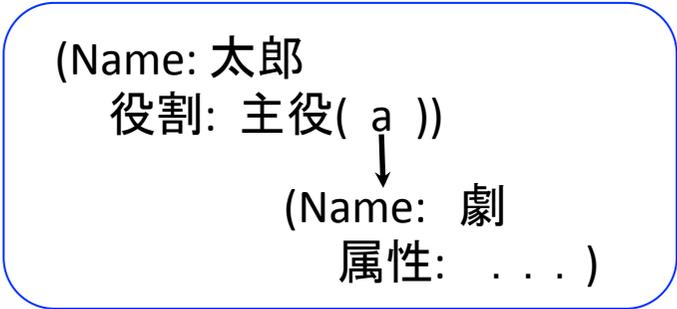
名詞には属性関係のような  
 指示しない名詞がある

「主役」vs「俳優」

太郎が その劇の主役だ

→ (主役(項: その劇))

飽和名詞: 建築家, 紳士, 車, 机  
 非飽和名詞: 作者, 委員長, 恋人



結局 オブジェクト-属性-属性値 の組ではないか(竹内)

意味ネットワーク

# 命題部分の処理に注目

- 命題部分を述語に書き換える
  - 述語の概念フレーム(語義)の異なり
    - 「彼は発表で上がった」「風呂から上がった」「煙が上がった」
    - 事例が整理されていて欲しい (辞書, 言語データ)
      - 言語資源
- アプローチ
  - 述語or単語の基本概念の事例を作っておく
  - 入力文でどの概念にマッチさせるか
    - 語義曖昧生解消 (奥村先生の本の5章参照)

# 言語資源

- 述語まわりの事例を含むもの
  - FrameNet (英語他) 無料
    - 意味概念をベースに動詞, 名詞, 副詞, 形容詞を関係づけ
  - WordNet (英語・日本語) 無料
    - 類語関係を記述. 名詞, 動詞他. 大規模
  - 日本語語彙体系 (日本語) (有料) (NTT)
    - 日本語の名詞, 動詞に関する意味関係を木構造で記述
  - EDR (日本語) (商用)
    - 日本語の名詞, 動詞, 形容詞, 副詞に関する巨大な概念辞書, 英語の翻訳辞書, 意味概念付与テキストデータ
  - 述語項構造シソーラス(岡山大)
    - 1.1万語, 2.3万事例
      - <http://pth.cl.cs.okayama-u.ac.jp/>
  - 日本語FrameNet
    - 慶応大学小原先生
  - IPAL(1986)
    - 情報処理振興事業協会

# 語義曖昧性解消の例

- 用例ベース

- 「拾う」の概念フレーム (語義というよりFrame)

- 1) 私が/動作主 財布を/対象 拾う (受け取り)
- 2) チームが/動作主 勝ちを/対象 拾う (取得)
- 3) タクシーを/対象 拾う (乗り降り)
- 4) 活字を拾う (抽出)

- 語義曖昧性解消

「太郎が 1000円を 拾った」はどの意味か?  
教科書 p.75 (用例に基づく手法)

# 用例に基づく語義曖昧性解消

- 手法
  - 「太郎が 1000円を 拾った」を用例1)~4)と比較して近いものを取り出す
- 用例との類似度

$$\begin{aligned} & \operatorname{argmax} \quad \operatorname{Sim}(\text{入力文}, \text{概念フレーム}) \\ & \text{概念フレーム} \\ & = \operatorname{sum} (\operatorname{sim}(\text{各要素}, \text{概念フレームの要素})) \end{aligned}$$

sim のやり方は色々  
教科書 p.75の手法

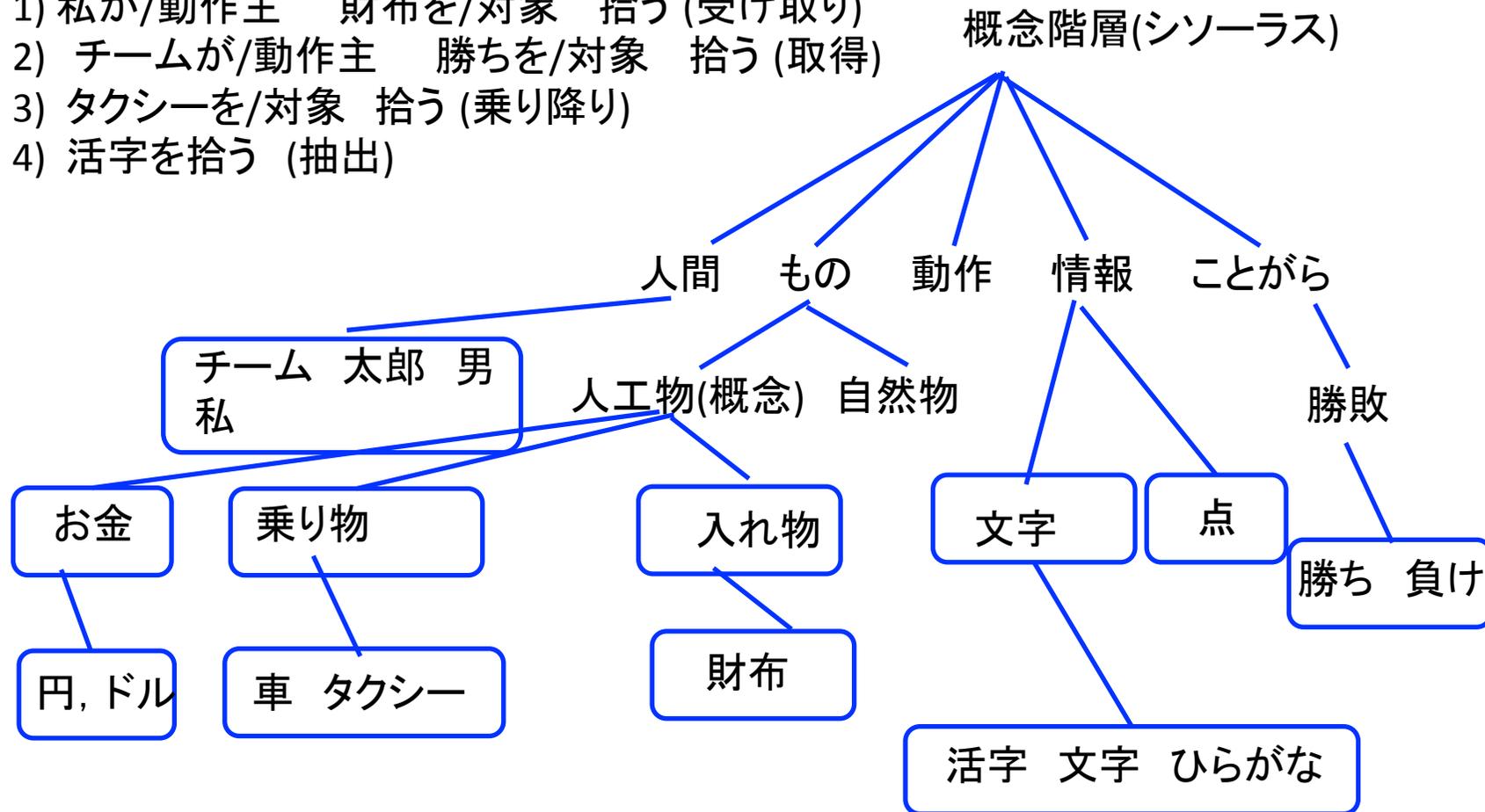
$$\begin{aligned} \operatorname{sim}(a, b) & \text{ 名詞のシソーラス(教科書)} \\ & = \text{共通の深さ} \times 2 / (a\text{の深さ} + b\text{の深さ}) \end{aligned}$$

別の方法

$$\begin{aligned} \operatorname{sim}(a, b) & = 2c_{xy} / (d_a + 1) (d_b + 1) \\ & \text{共通の深さ } c_{xy} \\ & d_a \text{ } a\text{の深さ}, d_b \text{ } b\text{の深さ} \end{aligned}$$

# 用例に基づく語義曖昧性解消

- 1) 私が/動作主 財布を/対象 拾う (受け取り)
- 2) チームが/動作主 勝ちを/対象 拾う (取得)
- 3) タクシーを/対象 拾う (乗り降り)
- 4) 活字を拾う (抽出)



# 類似度計算

- p.75の方法

1) 私が/動作主 財布を/対象 拾う (受け取り)

a) 「太郎が 1000円を 拾った」

事例の類似度計算

$$\text{Sim}(1,a) = \text{sim}(\text{私}, \text{太郎}) + \text{sim}(\text{財布}, \text{1000円})$$

$$= (1 \times 2) / (2+2) + (2+2)/(3+4)$$

$$= 1/2 + 4/7$$

# 練習

- 練習

- 「2) チームが/動作主 勝ちを/対象 捨う」と比較して  
1)の場合と比べてどちらが近いか答えなさい