

2018/12/19 16:30 修正

# 計算言語学<sub>10</sub>

## 意味解析

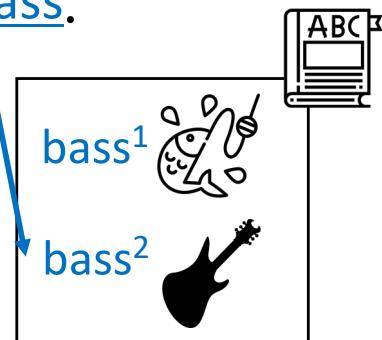
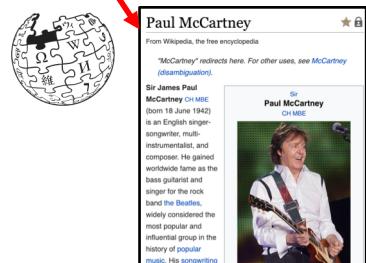
東京大学生産技術研究所  
吉永 直樹

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

# 意味解析

## 語義曖昧性解消 | Entity Linking

Paul played a black bass.



## 固有表現抽出 | supersense tagging

Ringo Star has joined the bands  
group

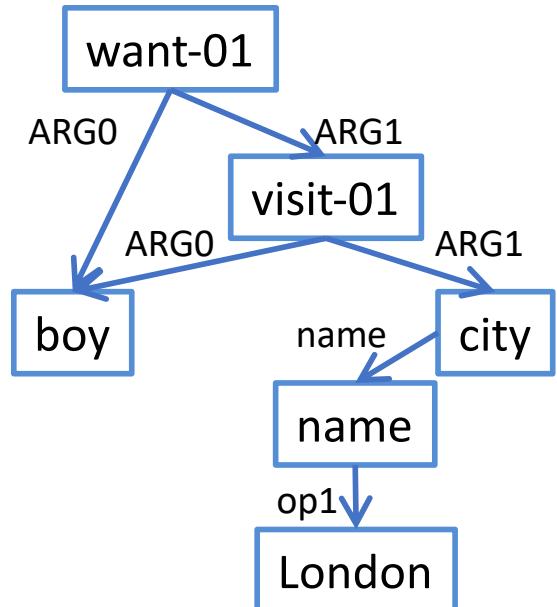
PERSON      Non-NE      Non-NE

## 意味役割付与 (Semantic Role Labeling; SRL)

The boy wants to visit London

ARG0      述語 (visit-01)      ARG1

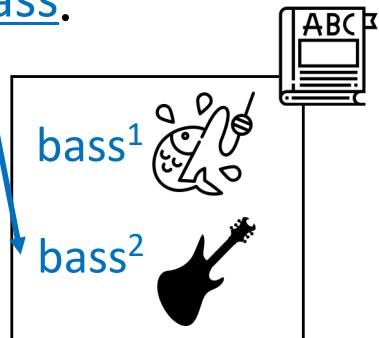
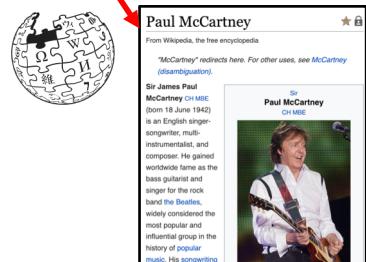
## Semantic Parsing



# 意味解析

## 語義曖昧性解消 | Entity Linking

Paul played a black bass.



## 固有表現抽出 | supersense tagging

Ringo Star has joined the bands  
group

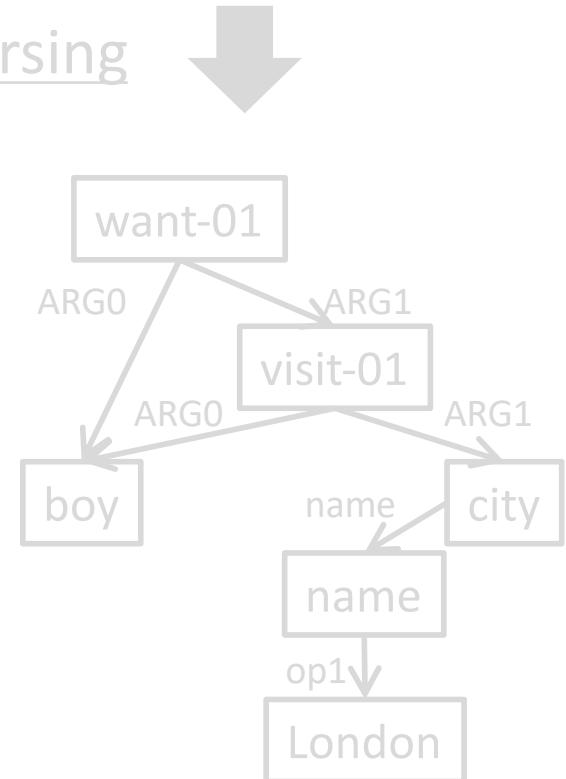
PERSON      Non-NE      Non-NE

## 意味役割付与 (Semantic Role Labeling; SRL)

The boy wants to visit London

ARG0      述語 (visit-01)      ARG1

## Semantic Parsing



# 単語・句レベルの意味解析

語義曖昧性解消 (1950s-)

固有表現認識 (1990s-)

粗粒度語義曖昧性解消 (2003-)

Entity Linking (2006-)

# 文中の単語・句の意味の曖昧性解消

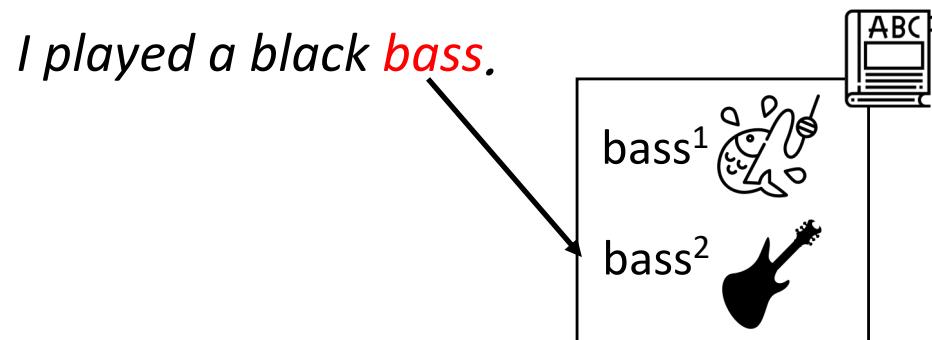
- 語義粒度の処理
    - Word Sense Disambiguation (WSD; 語義曖昧性解消)  
対象: 語句 (内容語)  
出力: 辞書中の語義
    - Entity Linking (エンティティリンク)  
対象: エンティティ  
出力: 知識ベースの項目
  - 意味カテゴリ粒度の処理
    - Supersense tagging (疎粒度語義曖昧性解消)  
対象: 一般名詞  
出力: 意味カテゴリ
    - Named Entity Recognition (NER; 固有表現抽出)  
対象: 固有表現  
出力: 意味カテゴリ(粗い)
- 
- |  |                 |  |                        |
|--|-----------------|--|------------------------|
|  | 粒度の細かい曖昧性解消     |  | 現実的問題設定 (高精度)          |
|  | 未知語・未知語義の扱い     |  | 同名・同意味カテゴリの語句の曖昧性解消は不可 |
|  | 語義毎に訓練例が必要(WSD) |  |                        |

# 文中の単語・句の意味の曖昧性解消

- 語義粒度の処理
    - Word Sense Disambiguation  
(WSD; 語義曖昧性解消)  
対象: 語句 (内容語)  
出力: 辞書中の語義
    - Entity Linking  
(エンティティリンク)
  - 意味カテゴリ粒度の処理
    - Supersense tagging  
(疎粒度語義曖昧性解消)  
対象: 一般名詞  
出力: 意味カテゴリ
    - Named Entity Recognition  
(NER; 固有表現抽出)  
対象: 固有表現  
出力: 意味カテゴリ(粗い)
- |  |                 |  |                            |
|--|-----------------|--|----------------------------|
|  | 粒度の細かい曖昧性解消     |  | 現実的問題設定(高精度)               |
|  | 未知語・未知語義の扱い     |  | 同名・同意味カテゴリの<br>語句は曖昧性解消できず |
|  | 語義毎に訓練例が必要(WSD) |  |                            |

# Word Sense Disambiguation (WSD; 語義曖昧性解消)

- 文中の単語(多義語)の意味を辞書から選択
  - 入力: 単語とその出現文脈 + 語義定義 (sense inventory)



- 対象とする語により2種類のタスクが存在
  - Lexical sample WSD: 特定の単語のみを対象  
教師あり学習が利用可能
  - All-words WSD: 辞書に含まれる全単語を対象  
教師あり学習は利用し辛い (十分な学習データがない)

# Word Sense Disambiguation: 語義をどう定義するか (WSD; 語義曖昧性解消)

- **Zeugma** (くびき語法)

異なる文脈で出現する単語の語義の同義性判定のための言語テスト

- Q: 以下の2つの文の *serve* の語義は区別すべきか ?

*Which of those flights **serve** breakfast?*

*Does Midwest Express **serve** Philadelphia?*

- A: 2文脈を繋げた文の意味が通らないので区別すべき

*? Does Midwest Express **serve** breakfast and Philadelphia?*

# Word Sense Disambiguation: データセット (英語) (WSD; 語義曖昧性解消)

- Lexical sample WSD
  - Line-hard-serve コーパス [Leacock+ 1993]: 3語, 各4000事例
  - Senseval-1,-2,-3 [Kilgarriff+ 1998, Kilgarriff+ 2001, Mihalcea+ 2004]: HECTOR 辞書 [Atkins 1993] 中の34, 73, 57語について, 多様な文書から事例収集
- All words WSD:
  - SemCor [Landes et al. 1998]: semantic concordance Brown コーパス中の234,000事例に WordNet の語義を付与
  - Senseval-2 [Palmer+ 2001]: WSJ の5000語中の内容語2081語に対して語義付与

# Word Sense Disambiguation: 教師有り手法 (WSD; 語義曖昧性解消)

- 機械学習の分類問題として定式化
  - 特徴量: 出現文脈中の語(品詞, 埋め込み) (BoW or 位置有)
  - Wikipedia の記事リンクから学習データ生成 [Mihalcea 2007]  
*It is danced in 3/4 time (like most waltzes), with the couple turning approx. 180 degrees every [[bar (music)|bar]].*

(曖昧性のない) Wikipedia記事
- 評価:
  - 本来はしきるべき応用で extrinsic な評価を行うべきだが, 効率性を重視して intrinsic な評価が行われることが多い
  - ベースライン: 最頻語義 (非常に強い)  
WordNet の語義は SemCor での語義頻度順例) Senseval-3 (all-words WSD):  
最頻語義 (66.0%), sota (70.6%) [Papandrea+ 2016]

# Word Sense Disambiguation: 弱教師有り手法 (1/2) (WSD; 語義曖昧性解消)

- Simplified Lesk アルゴリズム [Kilgarriff+ 2001]:
    - 辞書中の語釈文や用例と出現文脈の重複語数によるWSD
- 例) WordNet *signature*

bank <sup>1</sup>	Gloss (語釈文)	a financial institution that accepts <b>deposits</b> and channels the money into lending activities
	Examples (用例)	“he cashed a check at the bank”, “that bank holds the <b>mortgage</b> my home”
bank <sup>2</sup>	Gloss (語釈文)	sloping land (especially the slope beside a body of water)
	Examples (用例)	“they pulled the canoe up on the bank”, “he sat on the bank of the river and watched the currents”

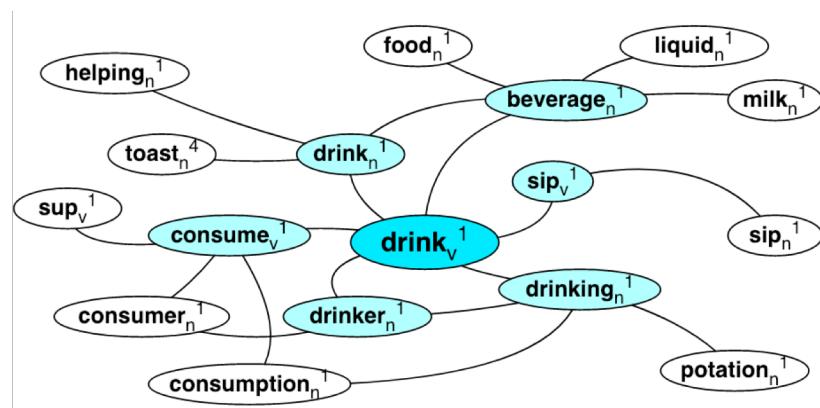
入力) *The **bank** can guarantee **deposits** will eventually cover future tuition costs because it invests in adjustable-rate **mortgage** securities*

- Corpus Lesk [Kilgarriff+ 2001]: 語義付与文を *signature* に利用
  - 重複語を IDF (各語の *signature* 中での逆文書頻度)で重み付け

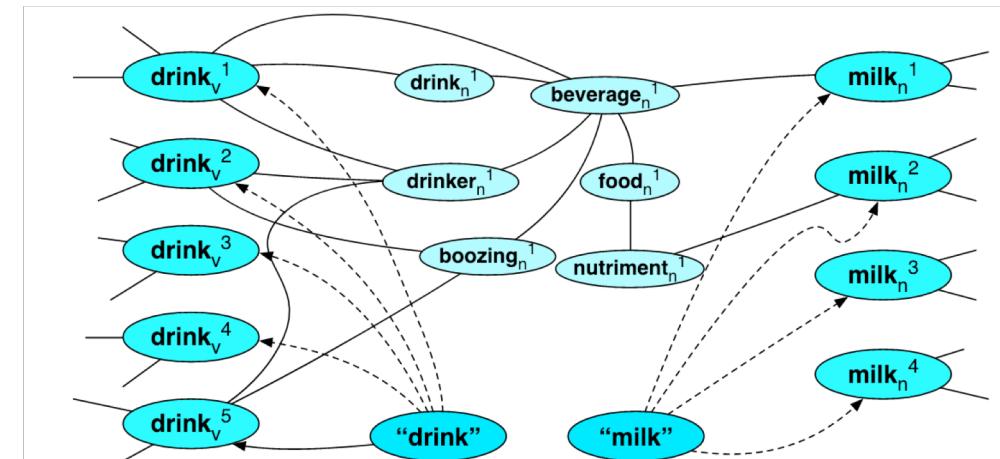
# Word Sense Disambiguation: 弱教師有り手法 (2/2) (WSD; 語義曖昧性解消)

- グラフに基づく手法 [Navigli+ 2010, Agirre+ 2014]
  - 対象語や出現文脈中の単語について、シソーラスで定義された語義間関係グラフを利用
  - ノードの次数やランダムウォーク (PageRank) などによりグラフにおける各ノードの中心性を計算して語義同定

WordNet 中の語義間関係グラフ

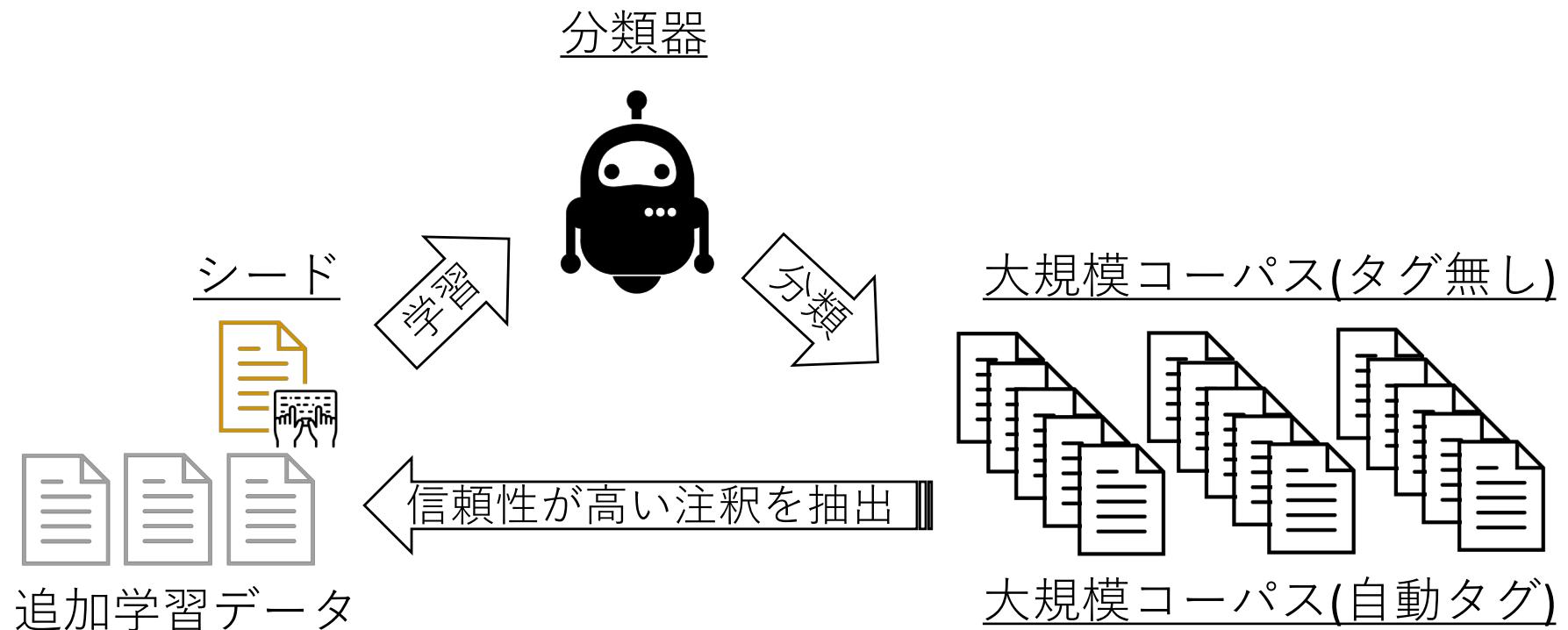


入力: *She drank some milk*



# Word Sense Disambiguation: 半教師有り手法 (1/2) (WSD; 語義曖昧性解消)

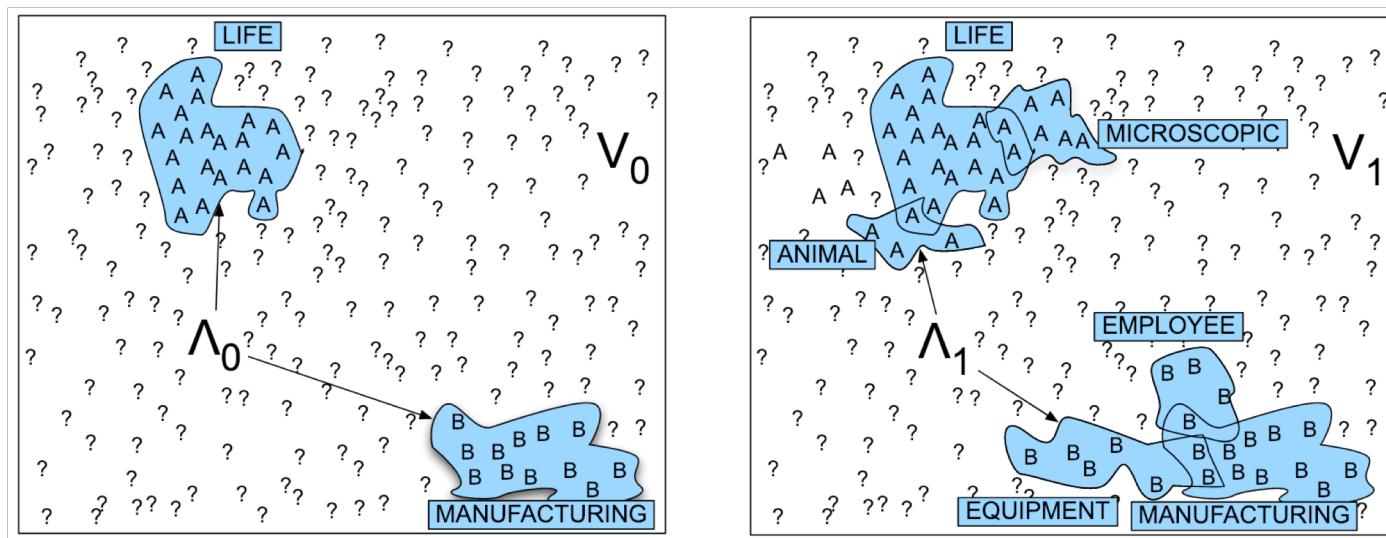
- **Bootstrapping** [Yarowsky 1995]:  
少数の訓練データ(シード)と生データから徐々に訓練データを拡張



# Word Sense Disambiguation: 半教師有り手法 (2/2) (WSD; 語義曖昧性解消)

- Bootstrapping [Yarowsky 1995]:  
少数の訓練データ(シード)と生データから徐々に訓練データを拡張
- 以下のヒューリスティクスをデータ生成に活用
  - One-sense-per-collocation: 連語中の語義は一意
  - One-sense-per-discourse [Gale+ 1993]: 文書中の語義は一意

*plant* の語義曖昧性解消における bootstrapping



# Word Sense Disambiguation: 語義推定 (WSD; 語義曖昧性解消)

- 問題: WSDは辞書にない未定義語(義)を扱えない
- タスク: 教師なし語義推定 [Schütze 1992, 1998]
  - 辞書を用いず、語義が同じ出現文脈を判別する
- アプローチ:
  - 単語の出現文脈をベクトルで表現してクラスタリングし得られた各クラスタ中心 (centroid) を語義と考える
  - テスト: 出現文脈のベクトル表現に最も近い語義を選択
- 評価:
  - 正解語義で出現文脈をクラスタリングした結果との一致

# 文中の単語・句の意味の曖昧性解消

## • 語義粒度の処理

- Word Sense Disambiguation  
(WSD; 語義曖昧性解消)  
対象: 語句 (内容語)  
出力: 辞書中の語義
- Entity Linking  
(エンティティリンク)  
対象: エンティティ  
出力: 知識ベースの項目

😊 粒度の細かい曖昧性解消

😢 未知語・未知語義の扱い

😢 語義毎に訓練例が必要(WSD)

## • 意味カテゴリ粒度の処理

- Supersense tagging  
(疎粒度語義曖昧性解消)  
対象: 一般名詞  
出力: 意味カテゴリ
- Named Entity Recognition  
(NER; 固有表現抽出)  
対象: 固有表現  
出力: 意味カテゴリ(粗い)

😊 現実的問題設定(高精度)

😢 同名・同意味カテゴリの語句は曖昧性解消できず

# Named Entity Recognition (1990s～) (NER; 固有表現抽出)

- 文中の固有表現 (Named entity)を自動認識し、  
そのタイプ(意味カテゴリ)の曖昧性解消を行う
  - 情報抽出 (Information Extraction)への応用を強く意識

MUC (Message Understanding Conference) における7タイプ  
ORGANIZATION, PERSON, LOCATION, DATE, TIME, MONEY, PERCENT

Ringo Star has joined the Beatles on Aug. 18th, 1962

PERSON                    ORGANIZATION                    TIME

- 分野依存の固有表現タイプ: 遺伝子, タンパク質 (BioNLP)
- 汎用の固有表現タイプの拡張の試み:  
拡張固有表現 (200種) [Sekine 2000], Freebase の利用 [Ritter+ 2011]

# Named Entity Recognition: 関根の拡張固有表現

## 名前 名前\_その他

### 人名

### 神名

#### 組織名 組織名\_その他

国際組織名  
公園組織名  
家系名  
民族名 民族名\_その他  
国籍名  
競技組織名 競技組織名\_その他  
プロ競技組織名  
競技リーグ名  
法人名 法人名\_その他  
企業名 / 企業グループ名  
政治的組織名 政治的組織名\_その他  
政府組織名 / 政党名 /  
内閣名 / 軍隊名

#### 施設名 施設名\_その他

施設部分名  
遺跡名 遺跡名\_その他  
古墳名  
GOE GOE\_その他  
公共機関名 / 学校名 / 研究機  
関名 / 取引所名 / 公園名 / 競  
技施設名 / 美術博物館名 / 動  
植物園名 / 遊園施設名 / 劇場  
名 / 神社寺名 / 停車場名 / 電  
車駅名 / 空港名 / 港名  
路線名 路線名\_その他  
電車路線名 / 道路名 / 運河名  
航路名 / トンネル名 / 橋名

### 地名 地名\_その他

温泉名  
GPE GPE\_その他  
市区町村名 / 郡名  
都道府県州名 / 国名  
地域名 地域名\_その他  
大陸地域名  
国内地域名  
地形名 地形名\_その他  
山地名 / 島名 / 河川名  
湖沼名 / 海洋名 / 湾名  
天体名 天体名\_その他  
恒星 / 惑星 / 星座  
アドレス アドレス\_その他  
郵便住所 / 電話番号  
電子メール / URL

### イベント名

催し物名 催し物名\_その他  
例祭名 / 競技会名  
会議名  
事件事故名 事件事故名\_その他  
戦争名  
自然現象名 自然災害名\_その他  
自然災害名  
地震名  
病気名 病気名\_その他  
動物病気名

### 製品名 製品名\_その他

材料名 / 衣類名 / 貨幣名 /  
医薬品名 / 武器名 / 株名 /  
賞名 / 黙章名 / 罪名 / 便名  
等級名 / キャラクター名 /  
識別番号  
乗り物名 乗り物名\_その他  
車名 / 列車名 / 飛行機名  
宇宙船名 / 船名  
食べ物名 食べ物名\_その他  
料理名  
芸術作品名 芸術作品名\_その他  
絵画名 / 番組名 / 映画名 /  
公演名 / 音楽名 / 文学名  
出版物名 出版物名\_その他  
新聞名 /雑誌名

#### 主義方式名 主義方式名\_その他

文化名 / 宗教名 / 学問名 /  
競技名 / 流派名 / 運動名  
理論名 / 政策計画名  
規則名 規則名\_その他  
条約名 / 法令名  
称号名 称号名\_その他  
地位職業名  
言語名 言語名\_その他  
国語名  
単位名 単位名\_その他  
通貨名

色名 色名\_その他  
自然色名

### 自然物名 自然物名\_その他

元素名  
化合物名  
鉱物名  
生物名 生物名\_その他  
真菌類名 / 軟体動物\_節足動物  
物名 / 昆虫類名 / 魚類名 両生  
類名 / 爬虫類名 / 爬虫類名 /  
鳥類名 / 哺乳類名 / 植物名  
生物部位名 生物部位名\_その他  
動物部位名 / 植物部位名

### 数値表現 数値表現\_その他

金額表現 / 株指標 / ポイン  
ト / 割合表現 / 倍数表現 /  
頻度表現 / 年齢 / 学齢 / 序  
数 / 順位表現 / 緯度経度  
寸法表現 寸法表現\_その他  
長さ / 面積 / 体積 / 重量 /  
速度 / 密度 / 溫度 / カロ  
リー / 震度 / マグニチュー  
ド / 個数 個数\_その他  
人数 / 組織数 / 場所数\_そ  
の他 / 国数 / 施設数 / 製品  
数 / イベント数 / 自然物数  
\_その他 / 動物数 / 植物数

### 時間表現

時刻表現 / 日付表現 / 曜日表  
現 / 時代表現 / 期間\_その他 /  
時刻期間 / 日数期間 / 週数期  
間 / 月数期間 / 年数期間

[<https://sites.google.com/site/extendednamedentity711/> より引用]

NER は基本的に教師あり学習で解くため、対応する  
言語資源があるか(高精度の分類が可能か)が重要

# Named Entity Recognition: 基本アプローチ (NER; 固有表現抽出)

- IOBタグ付けを経由した系列ラベリングにより、  
固有表現のチャンキングとクラス分類をモデル化
  - 条件付き確率場, 構造化パーセプトロン, BiLSTM-CRF etc.

Ringo has joined the Beatles on Aug. 18th, 1962

[IOB]	B-PER	O	O	B-ORG	I-ORG	O	B-TIME	I-TIME	I-TIME
[IOBES]	S-PER	O	O	B-ORG	E-ORG	O	B-TIME	I-TIME	E-TIME

- 特徴量:
  - 対象の語 + 前後の語 (prefix/suffix, 品詞, 埋め込み)
  - Gazetteer [Kazama+ 2007]: 地名辞書(広義: 知識ベース)との一致
- 評価: 固有表現 + ラベルの再現率, 精度,  $F_1$  値
  - SoTA: 93.09% (BiLSTM-CRF; [Akbik+ 2018]) (CoNLL03 dataset)

## 発展: Person Name Disambiguation for search results (Web 上の) 同姓同名問題の解消

- 問題: NER で付与される意味カテゴリは**粗粒度**
  - 同じ意味カテゴリで同名のエンティティがある場合、それらの間の曖昧性解消は行われない  
<https://ja.wikipedia.org/wiki/Category:すべての曖昧さ回避>
- タスク: Person Name Disambiguation [Bagga+ 1998, Mann+ 2003]
  - Web 上での人名検索における同姓同名問題という具体的な問題に焦点を絞り、問題を切り出して定式化
  - アプローチ: 教師なし学習(クラスタリング)

# 文中の単語・句の意味の曖昧性解消

- 語義粒度の処理
    - Word Sense Disambiguation  
(WSD; 語義曖昧性解消)  
対象: 語句(内容語)  
出力: 辞書中の語義
    - Entity Linking  
(エンティティリンク)
  - 意味カテゴリ粒度の処理
    - Supersense tagging  
(疎粒度語義曖昧性解消)  
対象: 一般名詞  
出力: 意味カテゴリ
    - Named Entity Recognition  
(NER; 固有表現抽出)  
対象: 固有表現  
出力: 意味カテゴリ(粗い)
- 
- |  |                 |  |                            |
|--|-----------------|--|----------------------------|
|  | 粒度の細かい曖昧性解消     |  | 現実的問題設定(高精度)               |
|  | 未知語・未知語義の扱い     |  | 同名・同意味カテゴリの<br>語句は曖昧性解消できず |
|  | 語義毎に訓練例が必要(WSD) |  |                            |

# Supersense tagging (Coarse-grained WSD) [Ciaramita+ 2003] (疎粒度語義曖昧性解消)

- 背景:
  - WSD は辞書にない未知語・未知語義を扱えない
  - NER は固有名詞のみが対象, 出力は粗い意味カテゴリ
- タスク: 一般名詞に語非依存の意味カテゴリを付与  
WordNet の意味体系の上位階層の語義 (**supersense**)

person	cognition	attribute	quantity	plant
communication	possession	object	motive	relation
artifact	location	process	animal	
act	substance	Tops	body	
group	state	phenomenon	feeling	
food	time	event	shape	

# 文中の単語・句の意味の曖昧性解消

## • 語義粒度の処理

- Word Sense Disambiguation  
(WSD; 語義曖昧性解消)

対象: 語句

出力: 辞書中の語義

- Entity Linking

(エンティティリンクング)

対象: エンティティ

出力: 知識ベースの項目



粒度の細かい曖昧性解消



未知語・未知語義の扱い



語義毎に訓練例が必要(WSD)

## • 意味カテゴリ粒度の処理

- Supersense tagging  
(疎粒度語義曖昧性解消)

対象: 一般名詞

出力: 意味カテゴリ

- Named Entity Recognition

(NER; 固有表現抽出)

対象: 固有表現

出力: 意味カテゴリ(粗い)



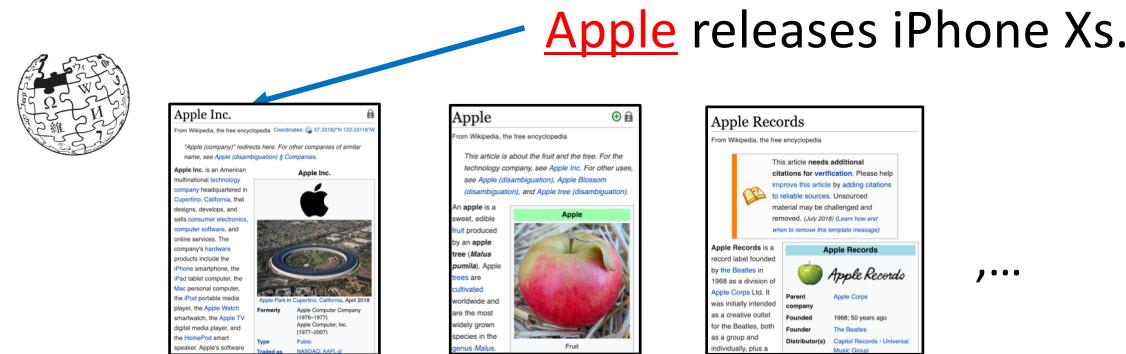
現実的問題設定(高精度)



同名・同意味カテゴリの  
語句は曖昧性解消できず

# Entity Linking [Bunescu+ 2006, Cucerzan+ 2007] (エンティティリンク)

- 背景:
  - NER は同名同意味カテゴリの固有表現を区別できない
  - 大規模知識ベース(Wikipedia)の出現 (記事=エンティティ)
- タスク: 文中のエンティティへの言及(mention)に対し  
言及が指す知識ベースの項目(記事)をリンク



- (入力文中の mention の同定 + ) リンク先の曖昧性解消
- 評価: 精度 (Micro: 全平均; Macro: 記事単位で平均)
  - Sota: 94.88% [Raiman+ 2018] (CoNLL-03 datasets)

# Entity Linking: 基本アプローチ

## 1. mention の同定:

NERや名詞句チャンキングで同定(所与とする設定が多い)

## 2. mention のリンク先(エンティティ)の絞り込み

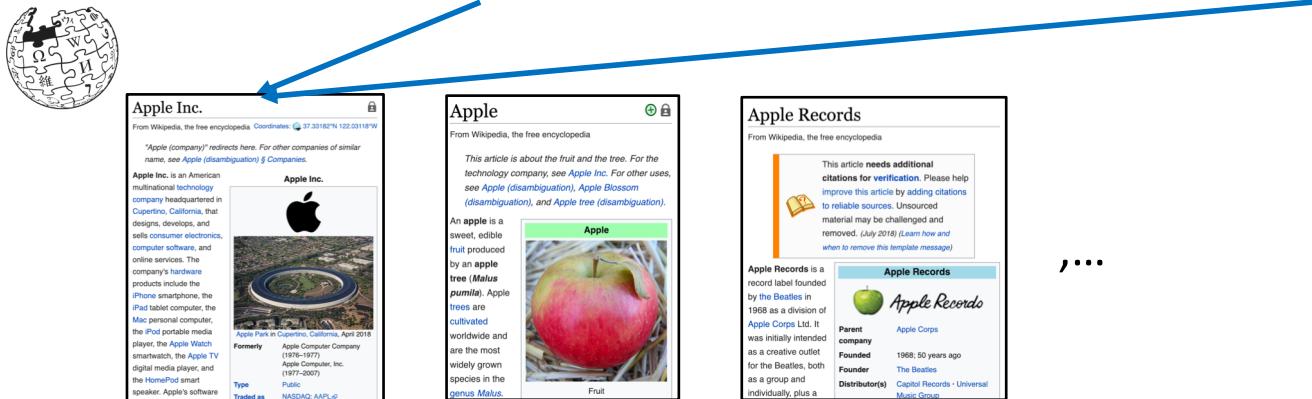
mention とエンティティ名(記事タイトル)との類似度を利用

## 3. エンティティ候補のランキング

2のエンティティの記事内容と出現文脈の類似度等を利用

注) リンクする記事がない mention もある

Apple releases iPhone Xs. Tim Cook, Apple CEO, started...



# Entity Linking:

局所的分類に基づくエンティティランギング

- 文脈-記事類似度 (mention)の出現文脈と候補記事の類似度)に基づく記事タイトルランギング [Bunescu+ 2006]

- 出現文脈と候補記事内容を tf-idf ベクトルで表現
- 記事リンクを利用した学習により次元を重み付け

*[[Tesla, Inc. | Tesla]] was founded in July 2003, ...*



[Apple Inc.](#)

From Wikipedia, the free encyclopedia. Coordinates: 37.3318°N 122.0318°W  
"Apple (company)" redirects here. For other companies of similar name, see [Apple \(disambiguation\)](#) & [Companies](#).

Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services. The company's core products include the iPhone smartphone, the iPad tablet computer, the Mac personal computer, the iPod portable media player, the Apple Watch smartwatch, the Apple TV digital media player, and the HomePod smart speaker. Apple's software

Formerly Apple Computer Company (1976–1997)  
Apple Computer, Inc. (1997–2007)

Type Public  
Traded as NASDAQ: AAPL, [AAPL\\_U](#)

[Apple](#)

From Wikipedia, the free encyclopedia

This article is about the fruit and the tree. For the technology company, see [Apple Inc.](#) For other uses, see [Apple \(disambiguation\)](#), [Apple Blossom \(disambiguation\)](#), and [Apple tree \(disambiguation\)](#).

An apple is a sweet, edible fruit produced by an apple tree (*Malus pumila*). Apple trees are cultivated worldwide and are the most widely grown species in the genus *Malus*.

 Apple

Fruit

[Apple Records](#)

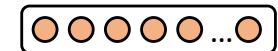
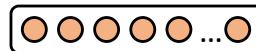
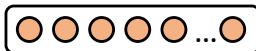
From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (July 2018) ([Learn how and when to remove this template message](#))

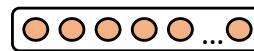
Apple Records is a record label founded by the Beatles in 1968 as a division of Apple Corps Ltd. It was initially intended as a creative outlet for the Beatles, both as a group and individually, plus a

 Apple Records

Parent company Apple Corps  
Founded 1968; 50 years ago  
Founder The Beatles  
Distributor(s) Capitol Records - Universal Music Group



Apple releases iPhone Xs.



コサイン類似度 +  
非零の次元の組合せ

$$\hat{t} = \operatorname{argmax}_t w^T \phi(m, t)$$

記事タイトルランギング

0.64 Apple Inc.

0.32 Apple Records

0.07 Apple

# Entity Linking: 大域的一貫性を考慮したエンティティランギング

- 局所的スコアに加え入力中の各mentionのリンク先の大域的一貫性を考慮して精度改善 [Ratinov+ 2011]

$$\hat{\Gamma} = \operatorname{argmax}_{\Gamma} \left[ \sum_{i=1}^N \phi(m_i, t_i) + \psi(\Gamma) \right]$$

$$= \operatorname{argmax}_{\Gamma} \left[ \sum_{i=1}^N \phi(m_i, t_i) + \sum_{t_j \in \Gamma'} \psi(t_i, t_j) \right]$$

記事タイトル間の一貫性

大域的一貫性を考慮する  
メンションのリンク先

Apple releases iPhone Xs. Tim Cook, Apple CEO, started...



**Apple Inc.**

From Wikipedia, the free encyclopedia. Coordinates: 37°33'18"N 122°03'18"W

"Apple (company)" redirects here. For other companies of similar name, see [Apple \(disambiguation\)](#) § Companies.

Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services. The company's hardware products include the iPhone smartphone, the iPad tablet computer, the Mac personal computer, the iPod portable media player, the Apple Watch smartwatch, the Apple TV digital media player, and the HomePod smart speaker. Apple's software

**Formerly** Apple Computer Company (1976–1997)  
Apple Computer, Inc. (1977–2007)

**Type** Public  
**Traded as** NASDAQ: AAPL.us

**Apple**

From Wikipedia, the free encyclopedia

This article is about the fruit and the tree. For the technology company, see [Apple Inc.](#) For other uses, see [Apple \(disambiguation\)](#), [Apple Blossom \(disambiguation\)](#), and [Apple tree \(disambiguation\)](#).

An apple is a sweet, edible fruit produced by an apple tree (*Malus pumila*). Apple trees are cultivated worldwide and are the most widely grown species in the genus *Malus*.

**Apple Records**

From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (July 2018) ([Learn how](#) and [when to remove this template message](#))

**Apple Records** is a record label founded by the Beatles in 1968 as a division of [Apple Corps Ltd](#). It was initially intended as a creative outlet for the Beatles, both as a group and individually, plus a

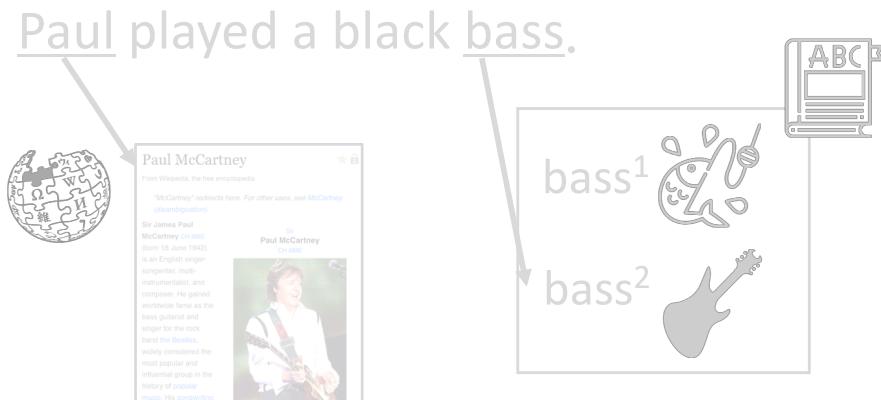
<b>Parent company</b>	<a href="#">Apple Corps</a>
<b>Founded</b>	1968; 50 years ago
<b>Founder</b>	The Beatles
<b>Distributor(s)</b>	Capitol Records - Universal Music Group

# 文の意味解析

Semantic Role Labeling (SRL; 意味役割付与)  
Semantic Parsing

# 意味解析

## 語義曖昧性解消 | Entity Linking



## 固有表現抽出 | supersense tagging

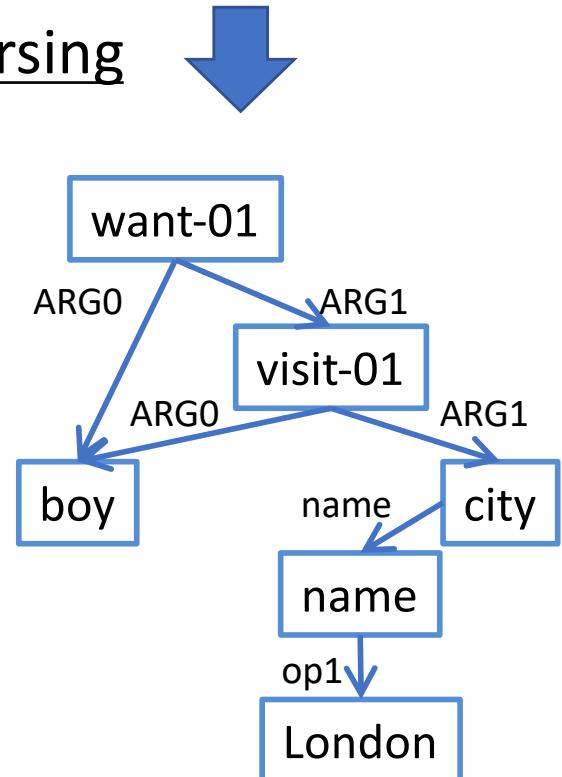
Ringo Star has joined the bands  
PERSON      Non-NE      Non-NE

## 意味役割付与 (Semantic Role Labeling; SRL)

The boy wants to visit London

ARG0      述語 (visit-01)      ARG1

## Semantic Parsing



# 意味役割付与 (Semantic Role Labeling; SRL)

- 文中の述語(動詞)と項(名詞)との間の意味的関係の分類に焦点を当てた意味解析

- *Kārakas*: Pāṇini による先駆的研究 (対象: 古 Sanskrit 語)  
*agent, instrument, destination etc.*

XYZ corporation **bought** the stock.

They **sold** the stock to XYZ corporation.

The stock was **bought** by XYZ corporation.

The **purchase** of the stock by XYZ corporation...

The stock **purchase** by XYZ corporation...

どれも「XYZ 株式会社が株を買ったイベント」を記述

- **意味役割**: 動詞に対して項が有する意味的関係
  - BUYER < AGENT < PROTO-AGENT (粒度は定義次第)

# 意味役割 (Semantic Role, Thematic Role): [Fillmore 1968, Gruber 1965]

- thematic role: 汎述語(イベント)的な抽象的意味役割

Thematic role	定義
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

$\exists e, y \text{Breaking}(e) \wedge \text{Breaker}(e, \text{Sasha}) \wedge \text{BrokenThing}(e, y) \wedge \text{Window}(y)$

$\exists e, y \text{Opening}(e) \wedge \text{Opener}(e, \text{Pat}) \wedge \text{OpenedThing}(e, y) \wedge \text{Door}(y)$

# 意味役割 (Semantic Role, Thematic Role): 例 [Fillmore 1968, Gruber 1965]

- thematic role: 汎述語(イベント)的な抽象的意味役割

Thematic role	例
AGENT	<i>The waiter <u>spilled</u> the soup</i>
EXPERIENCER	<i>John <u>has</u> a headache.</i>
FORCE	<i>The wind <u>blows</u> debris from the mall into our yards.</i>
THEME	<i>Only after Benjamin Franklin <u>broke</u> the ice...</i>
RESULT	<i>The city <u>built</u> a regulation-size baseball diamond...</i>
CONTENT	<i>Mona <u>asked</u> "You met Mary Ann at a supermarket?"</i>
INSTRUMENT	<i>He got catfish, <u>stunning</u> them with a <u>shocking</u> device</i>
BENEFICIARY	<i>Whenever Ann <u>makes</u> hotel reservations for her boss...</i>
SOURCE	<i>I <u>flew</u> in from <u>Boston</u>.</i>
GOAL	<i>I <u>drove</u> to <u>Portland</u>.</i>

# Diathesis alternation

- 意味役割を考えることで、述語に関する構文交替(diathesis alternation)を捨象できる

John<sub>AGENT</sub> broke the window<sub>THEME</sub>.

John<sub>AGENT</sub> broke the window<sub>THEME</sub> with a rock<sub>INSTRUMENT</sub>

The rock<sub>INSTRUMENT</sub> broke the window<sub>THEME</sub>.

The window<sub>THEME</sub> broke.

The window<sub>THEME</sub> was broken by John<sub>AGENT</sub>.

- 述語が取る項の意味役割の集合を  
格フレーム (thematic grid; θ-grid) と呼ぶ
- VerbNet [Kipper+ 2000]:  
動詞の構文交替パターンに基づく意味クラス [Levin+ 1993]

# 実用的な意味役割はどう設計すれば良いか？

- 前述の thematic role (单一の意味役割集合) は  
**帶に短し櫛に長し**
  - 研究対象に応じて、より細かく分ける必要が出てくる  
例) Levin+ (2005) による instrument 格の分析

The cook opened the jar with **the new gadget.** *intermediary instrument*  
The **new gadget** opened the jar.

Shelly ate the sliced banana with **a fork.** *enabling instrument*  
\***The fork** ate the sliced banana.
- 意味役割の階層化
  - 汎化意味役割: Proto-Agent, Proto-Patient [Dowty 1991]
  - 特定の述語や述語グループに特有の意味役割

# 意味役割付与: コーパス (1/2)

## PropBank

<https://propbank.github.io/>

- Proposition Bank (PropBank) [Palmer+ 2005]
  - Penn TreeBank (WSJ) (4万文)に対し動詞の意味役割を付与
  - 意味役割: ARG0-5 (語義に基づくフレーム依存),  
AM-\* (語義非依存; AM-LOC, AM-TMP など14種)  
Arg0 = PROTO-AGENT, Arg1 = PROTO-PATIENT 相当

The group Arg0 agreed it wouldn't make an offerArg1.

Usually AM-TMP John Arg0 agrees with MaryArg2 on everythingArg1.

注) 異なるフレームの意味役割  
間の関係は明示されない

agree.01  
Arg0: Agreer  
Arg1: Proposition  
Arg2: Other entity agreeing

- NomBank [Meyers+ 2004]: 述語性名詞に対する意味役割付与

# 意味役割付与: コーパス (2/2)

## FrameNet

<https://framenet.icsi.berkeley.edu/fndrupal/>

- FrameNet [Baker+ 1998, Fillmore+ 2003, Fillmore+ 2009, Ruppenhofer+ 2016]
  - フレーム意味論 [Fillmore 1985] に基づき, BNCコーパス +  $\alpha$  に汎述語的フレームに基づく意味役割を付与 (>20万文)  
例) **change\_position\_on\_a\_scale** frame: This frame consists of words that indicate the change of an Item's position on a scale (**the Attribute**) from a starting point (**Initial\_value**) to an end point (**Final\_value**).  
  
Oil<sub>ITEM</sub> **rose** in price<sub>ATTRIBUTE</sub> by 2%<sub>DIFFERENCE</sub>.  
It<sub>ITEM</sub> has **increased** to having them 1 day a month<sub>FINAL\_STATE</sub>.  
Microsoft shares<sub>ITEM</sub> **fell** to 7 5/8<sub>FINAL\_VALUE</sub>.  
Colon cancer incidence<sub>ITEM</sub> **fell** by 50%<sub>DIFFERENCE</sub> among men<sub>GROUP</sub>.  
a steady **increase** from 9.5<sub>INITIAL\_VALUE</sub> to 14.3<sub>FINAL\_VALUE</sub> in dividends<sub>ITEM</sub>
- 意味役割(フレーム要素): **core** (Arg0-5相当), non-core
- フレーム間の階層関係を定義

# 意味役割付与: 基本的なアプローチ

- タスク: 文中から述語とそのフレームを同定し,  
項となる構成素を同定し意味役割を付与
- アプローチ: 構文木の各ノードの意味役割を分類する多クラス分類問題(教師あり学習)
  - 特徴量: 述語の動詞(と下位範疇化フレーム),  
ノード(構成素)の品詞/主辞(品詞)/態/位置,  
述語からノードへの構文木上のパス
  - 意味役割間の依存関係を考慮 (Viterbi, re-ranking, ILPなど)
- 評価: 項の構成素 + 意味役割の再現率, 精度,  $F_1$ 
  - Sota: [Strubell+ 2018] 86.9% (CoNLL-2005 [Carreas+ 2005]),  
83.38% (CoNLL-2012 [Pradhan+ 2013])

# 発展: Selectional restriction / preference

- **Selectional restriction (選択制限):**  
述語が項に対し要求する意味的な制約をモデル化
  - 意味解析などの利用を想定

$$\exists e, x, y \underline{Eating(e)} \wedge Agent(e, x) \wedge \underline{Theme(e, y)}$$

「食べることができるもの」が来て欲しい

$$\exists e, x, y \underline{Eating(e)} \wedge Agent(e, x) \wedge \underline{Theme(e, y)} \wedge \underline{EdibleThing(y)}$$

WordNet の systet による表現  
{ food, nutorient }

- **Selectional preference (選択選好):**  
選択制限を緩めて確率的に扱うアプローチ(省略)

# Semantic parsing

- 文を意味表現に変換するタスクの総称

cf. semantic role labeling (SRL) = *shallow semantic parsing*

汎用	<ul style="list-style-type: none"><li>一階述語論理 (First-order logic): 前回講義参照</li></ul>
応用	<ul style="list-style-type: none"><li>Abstract Meaning Representation (AMR) [Barnarescu+ 2013]:</li></ul>
特化	<ul style="list-style-type: none"><li>Dependency-based compositional semantics (DCS) [Liang+ 2011]</li><li>SQL クエリ [Iyer+ 2017]</li><li>プログラムコード(やその抽象構文木) [Ling+ 2016, Yin+ 2017]</li><li>Alexa Meaning Representation Language (AlexaMRL) [Fan+ 2017]</li></ul>
・応用:	<ul style="list-style-type: none"><li>知識ベース (DB) に基づく質問応答システム</li><li>ロボット・知的エージェントの操作・ [Artzi+ 2013]</li><li>自然言語による推論</li></ul>

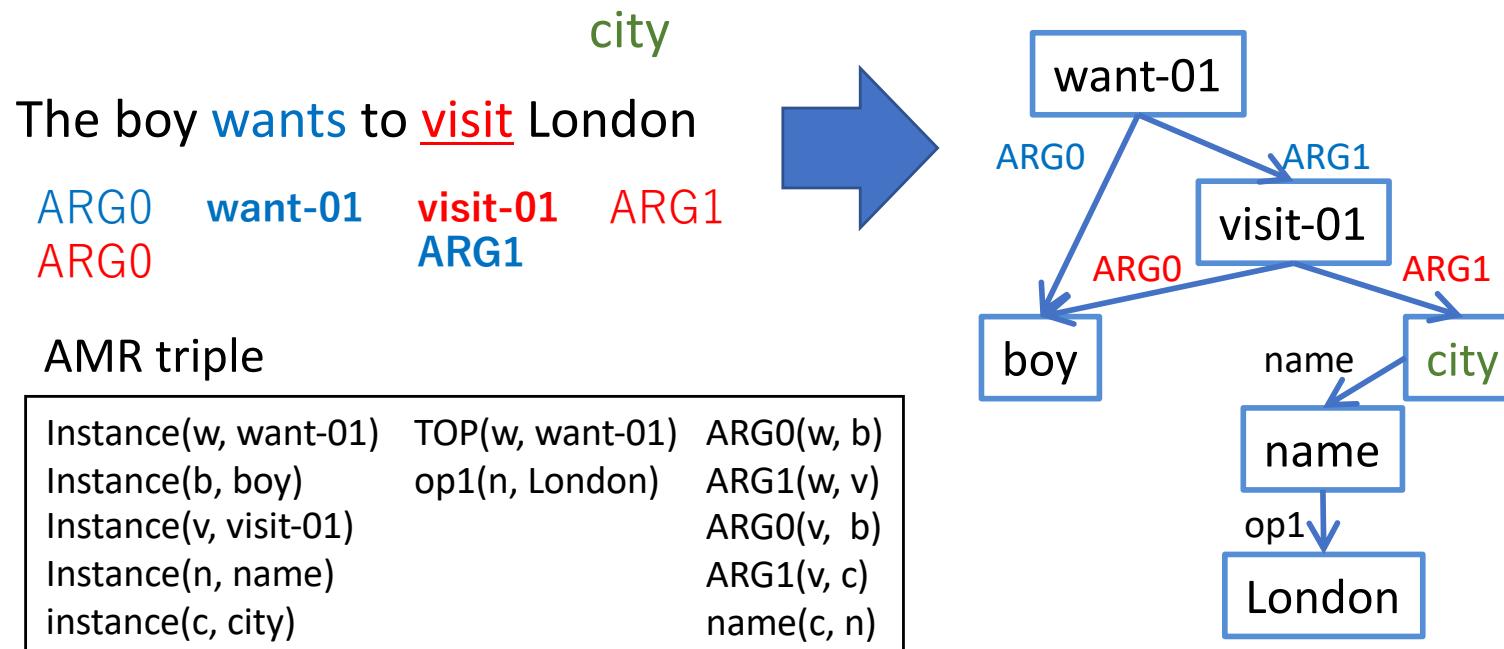
# Abstract Meaning Representation [Barnarescu+ 2013]

- 様々な既定の意味解析タスクの結果を統合可能な英語のための抽象的意味表現(グラフ表現)
  - 固有表現抽出, 意味役割付与, 共参照解析など



# Abstract Meaning Representation [Barnarescu+ 2013]

- 様々な既定の意味解析タスクの結果を統合可能な英語のための抽象的意味表現(グラフ表現)  
意味役割付与, 固有表現抽出, 共参照解析など
  - PropBank のフレーム意味論に強く依存
  - タグ付けの高速化ため, 文の意味を完全には保持しない  
例) 屈折, 全称量化子などは取り扱わない



# Abstract Meaning Representation [Barnarescu+ 2013]: 基本的アプローチ

- 構文解析ベースの手法
  - 依存構造器からの変換 [Wang+ 2015]
  - 遷移型構文解析の拡張による直接生成 [Ballesteros+ 2012]
- 機械翻訳 (Machine Translation; MT) ベースの手法
  - String-to-tree syntax-based statistical MT [Pust+ 2015]
  - Seq2seq (Neural MT): linearized AMRを出力 [Barzdins+ 2016]
- 評価尺度: smatch [Cai+ 2013]
  - AMR triple を単位として正解 AMR との一致度を計算
  - 変数については  $F_1$  値が最大となるマッチングを考慮

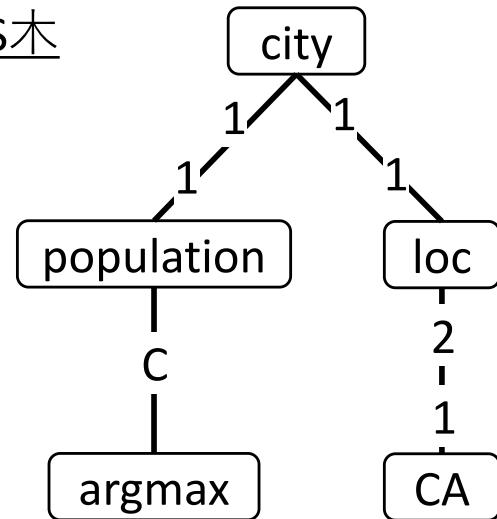
# Dependency-based Compositional Semantics (DCS)

[Liang+ 2011]

- データベースへの問い合わせに焦点を当てて設計された意味表現
  - 意味表現は実際の問い合わせと結果から学習 (省略)

*Q: What is the most populous city in California?*

DCS木



*A: Los Angels*

データベース

city	state
San Francisico	Alabama
Chicago	Alasca
Boston	Arizona
...	...

loc	
Mount Shasta	California
San Francisco	California
Boston	Massachusetts
...	...

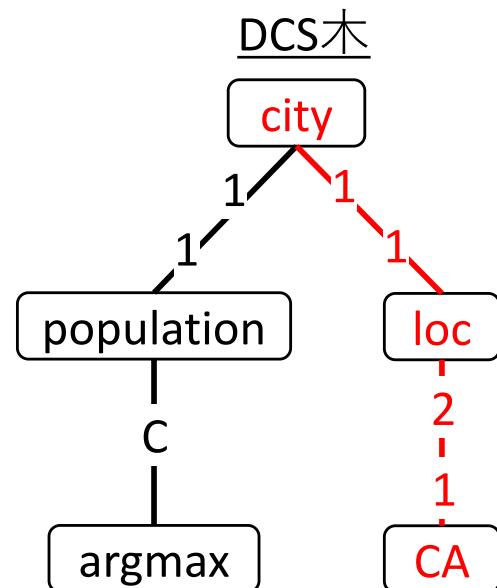
CA	California
	California

# Dependency-based Compositional Semantics (DCS)

[Liang+ 2011]

- データベースへの問い合わせに焦点を当てて設計された意味表現
  - 意味表現は実際の問い合わせと結果から学習 (省略)

*Q: What is the most populous city in California?*



Constraints

$c \in \text{city}$

$l_1 \in \text{loc}$

$l_2 = s_1$

$s \in \text{state}$

データベース

city	state
San Francisco	Alabama
Chicago	Alaska
Boston	Arizona
...	...

loc	
Mount Shasta	California
San Francisco	California
Boston	Massachusetts
...	...

CA	
	California

*A: Los Angeles*

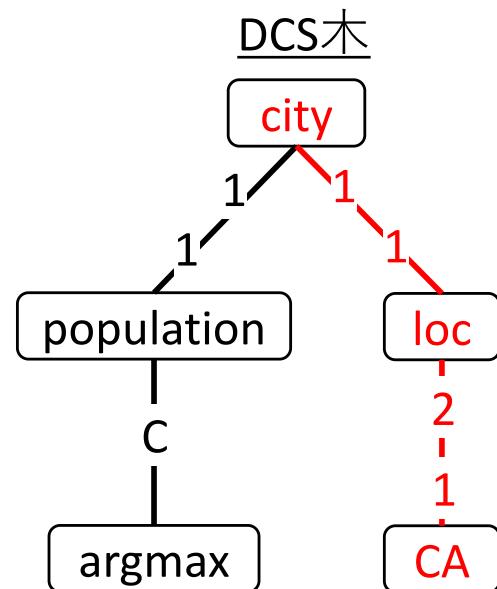
DCS木 = 制約充足問題

# Dependency-based Compositional Semantics (DCS)

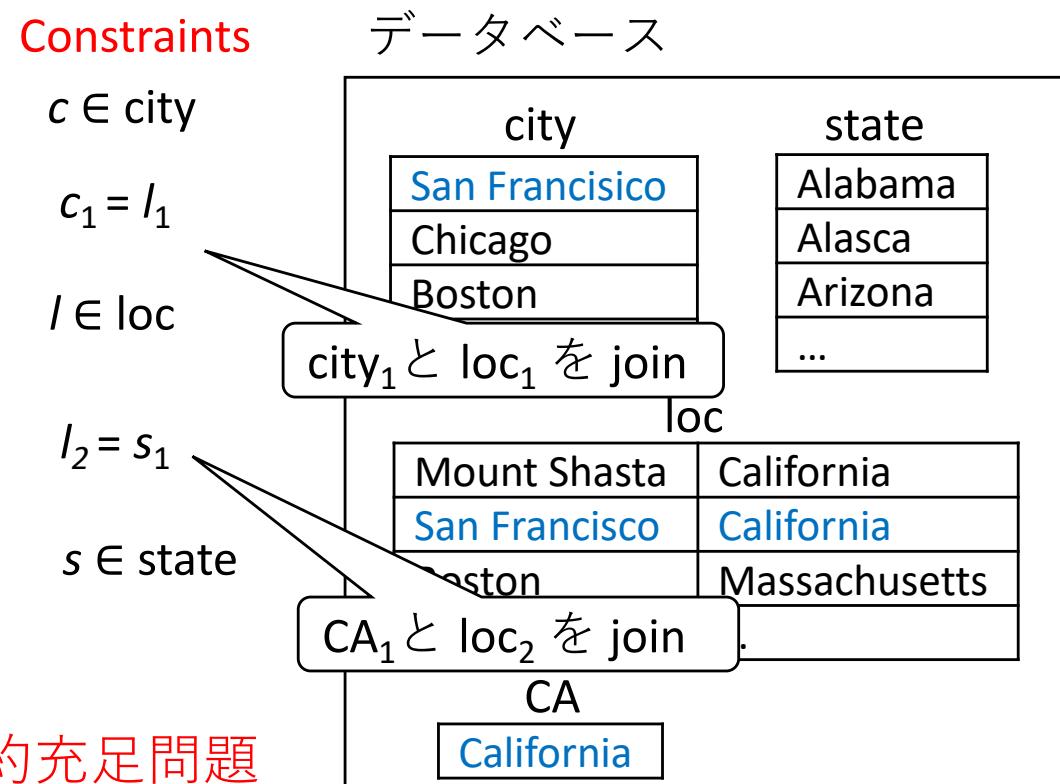
[Liang+ 2011]

- データベースへの問い合わせに焦点を当てて設計された意味表現
  - 意味表現は実際の問い合わせと結果から学習 (省略)

*Q: What is the most populous city in California?*



*A: Los Angels*



# まとめ

- 単語や句レベルの意味解析
  - 語義曖昧性解消 (Word Sense Disambiguation; WSD)
  - 固有表現抽出 (Named Entity Recognition; NER)
  - 粗粒度語義曖昧性解消 (coars-grained WSD; supersense tagging)
  - エンティティリンク (Entity Linking)
- 文の意味解析
  - 意味役割付与 (semantic role labeling)
  - Semantic Parsing (AMR, DCS)