

# 言語解析論

講師 竹内孔一

# 本日の内容

- 言語処理の応用
  - 質問応答
    - 知識の収集
    - 言い換え処理
    - 最新の話

# 質問応答システムの要件

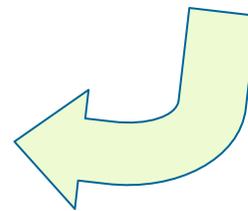
- 知識の構築
  - 既にあるテキストデータから自動での収集
  - > 未知の知識をなるべく人手をかけずに**獲得**
- 短い質問文
  - 前後の**背景・意図**を分析
    - 文部科学大臣の名前は？（「今の」や「日本の」が省略）
  - 複雑な**背景の解析**
    - 条件部分がほとんど
    - 「演習室でChaSenをインストールしたのですが動きません。どうすればよいでしょうか？」

# 知識の収集

- Webからのテキストデータの獲得
  - 文を係り受け解析して関係を整理
    - 意味関係推定も行う

..最近, 変なメールを受け取ることが多くなってきた..

Agent: 私  
Event\_obj: 変なメール  
Event: 受け取る  
Event\_time: 最近  
State: 多くなってきた



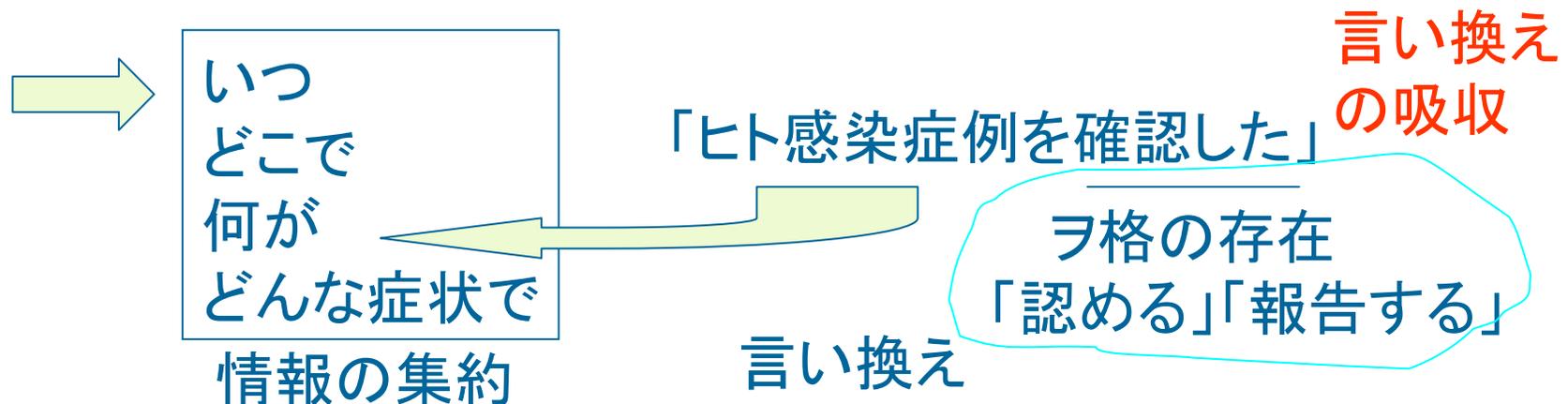
データベース化

項構造レベルの整理 + 固有表現抽出

# 知識の収集

- (例) バイオ関連ニュース

本日<NAME ci="LOCATION">〇〇国</NAME>  
<NAME ci="ORGANIZATION">保健省</NAME>は、  
<NAME ci="VIRUS">H5亜型鳥インフルエンザウイルス  
</NAME>の<NAME ci="PERSON" case="true"  
number="many">ヒト感染症例</NAME>をさらに  
<NAME ci="PERSON" case="true" number="many">  
5例</NAME> **確認した**



# 知識収集における問題

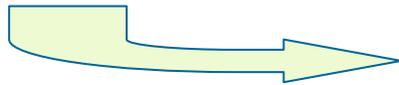
- 固有表現抽出
  - 確率統計モデルによる
  - 人や物の名前, 組織, 数 (分類)
- 項構造の同定
  - 動詞と結びつく名詞
  - [X(動作主)がY(対象)を確認する]
- 言い換え(語義によって異なる)
  - 「Xが症例を確認する/認知する」
  - 「先生が生徒のやる気を確認する/尋ねる」

WordNet, FrameNet (言語資源) 日本語ではまだ構築中

# 知識の収集

- 意味役割による情報抽出

×××は10日、〇〇〇国△△△県〇〇〇〇地域で先月死亡した農民の男性(62)が高病原性鳥インフルエンザに感染していたことを明らかにした。



誰: 農民の男性  
症状: 鳥インフルエンザ  
患者現状: 死亡

意味役割

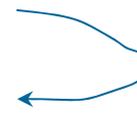
死亡した(V) ← 農民の男性 (Agent)



農民の男性 (Experiencer)  
鳥インフルエンザ (Theme)  
感染 (V)

情報の整理

属性が  
変わる



# 知識の収集-その他

- 手法2
  - FAQを探し質問にぴったりの回答を提示する
- 手法3
  - 検索システムに対する検索式として質問文を分解して答えを探す

# 練習21

- 「演習室でChaSenをインストールする方法は?」という質問に答えられる質問応答システムを考える

FAQ検索タイプのシステムはどのように設計できるか?

Webから知識獲得を行い質問に答えるシステムではどう設計する?

# 質問応答システム

- 本質的な問題

- 知識の基(限定)

- 言葉で書かれたデータ(Webなど)
- 整理されたデータベース
- 言葉になっていない規則

色, 音, 映像, 味, 感覚, 物理的な法則, 観念

- 質問文の解釈

- 文章の持つ意味の分解

- マッチング(推論手段)

- 質問文と知識とのマッチング

これらの先に会話処理

# 言い換え技術

- 言葉の本質的な問題

- 表現の同様なものをまとめる

例) 広辞苑第五版が11月11日に発売される

国語辞典として有名な広辞苑の第五版が11月11日に刊行される

- 質問

- 広辞苑第五版はいつ発売された?

- > 広辞苑第五版は<X>発売された?

時制  
複合語  
類義語

2002年頃からの新たな技術

# 言い換えの種類

- 構文と辞書に関連する言い換え
  - スタートする→始まる(類語)
  - 高田さんの造語だ→作った言葉だ(名詞句展開)
  - 彼は猫を2匹飼っている
    - 猫を2匹彼は飼っている(構文構造)
  - ディープブルーの軍門に降った→ディープブルーに負けた(イディオムから一般語)
  - 岡山大学→岡大(省略語)

# 言い換え

- 推論的な言い換え

- トリノは人口220万人の都市 → トリノの人口は220万人（「都市」を省略）
- フランスではフランス語が公用語である → フランス人はフランス語を話す（かなり推論が必要）

- 特徴

- 文における事実事象を変換していく
- (注意) 情報は必ずしも等価ではない

## 練習22

- 以下の知識と質問文がある。言い換えによりマッチングする様子を示せ

(知識) 明石海峡大橋は全長3911メートルである

(質問) 明石大橋は何メートルあるの？