

# 意味パターン翻訳システム

鳥取大学  
徳久 雅人

2009.10/31-11/1  
機械翻訳シンポジウム「機械翻訳研究の課題と新方式の展望」  
主催：福岡大学理学部柴田勝征研究室

1

## 1. はじめに

### ■ 本発表の目的

- 言語表現と意味と応用の考え方
  - 意味  
言語表現と概念の対応
  - パターン  
言語表現の抽象的概念を捕獲する道具
  - 翻訳  
同じ/近い概念と対応する2つの言語表現の変換
- パターン翻訳のデモ

2

## 2. 意味

意味＝言語表現と概念の対応

### ■ 言語表現とは

- 単語, 句, 節, 文, 文章
- 音素, 文字の並び
- 社会的な約束のもとでの並び

3

## 言語表現の生成の過程

(三浦:言語過程説)



4

## 言語表現の理解の過程

レーシングカーだ。

言語表現の表す概念に基づき、聞き手が認識する。

① 言語表現



② 認識



言語表現の表す概念は社会的な約束がある。(100%一致する訳ではない)

言語表現の表す概念から連想して深い認識をする。

5

## 意味処理

(池原2001)

① 言語表現

意味解析

言語表現を約束通りに解釈すること (約束の範囲内の処理)

② 浅い認識

意味理解

追加的な推論

②' 深い認識

文脈との結びつき  
既知のこととの比較  
連想

6

### 3. パターン

#### パターン

= 言語表現の表す抽象的概念を捕獲する道具

#### ■ 言語表現の表す概念

- 言語表現の一部でも変更すると、表される具体的な概念は、変化する。
  - 父さんに洋服を買ってもらった。
  - パパに洋服を買ってもらった。

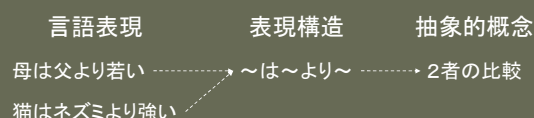
すべての言語表現を覚える必要がある!?

7

### 言語表現の構造と抽象的概念

#### ■ 言語表現の構造

- 言語表現の組み合わせ方
- 人々の間で共通の解釈がある



抽象的概念を表す必須の部分=非線形要素  
必須ではない部分=線形要素

8

### できるだけ正確に抽象化するために

- 文法的・意味的に条件を付与=パターン化

#### 2者の比較?

- 条件なし: ～は～より～。
  - 川は山より流れる。
- 文法的条件: M1 は N2 より AJ3。
  - 気温は鳥取より高い。
- 意味的条件: M1 (具体物) は N2 (具体物) より AJ3。
  - 鼻は像より長い。

100% 正確にという訳にはいかない

- 解釈の候補を出す◎ 解釈の非候補を排除×

9

### パターンに基づく意味処理

1. 非常にたくさんのパターンと付属情報を準備

<b>パターン</b>	<b>付属情報</b>
P1: ～は～より～	2者の比較と解釈
P2: ～は～より～	比較観点と対象と解釈
P3: ～は～より～	事象の起点と解釈
---	---

2. 入力文に適合するパターンを全て出力
  - たくさんの解釈の候補を出す
  - 曖昧なものは曖昧なままで処理を進行
3. 最終段階で曖昧性を解消

10

### 4. 機械翻訳

#### 機械翻訳

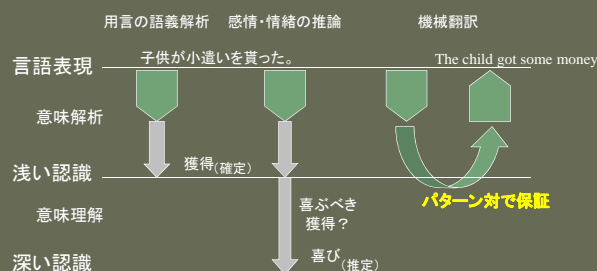
= 同じ/近い概念と対応する2つの言語表現の変換

#### ■ 要求される事項

- 源言語の表現に対するパターン (解析目的)
- 目的言語の表現に対するパターン (生成目的)
- 両パターンが同じ/近い抽象的概念を表す保証
- 候補の絞り込み

11

### 意味処理のレベルと機械翻訳



12

## パターン知識の要件

- (a) 文型パターンの記述
- (b) 抽象的概念の捉え方
- (c) 応用のための情報の付与

13

## 日本語語彙大系

- 文型パターンの記述  
N1がN2をN3から貰う  
N1(主体) N2(\*) N3(主体)
- 抽象的概念  
用言意味属性: 所有的移動
- 応用のための情報  
N1 get N2 from N3

14

## (参考)情緒推定用のパターン

- 文型パターンの記述 (1万パターン)  
N1がN2をN3から貰う  
N1(主体) N2(\*) N3(主体)
- 抽象的概念  
用言意味属性: 所有的移動
- 応用のための情報  
前提条件: 目標実現・近(N1, N2) 対人関係・近(N1, N3)  
情緒名: 喜び 好ましい  
情緒原因: 獲得 厚遇  
情緒主: M1 M1  
情緒対象: N2 N3

15

## (参考)対話感情文末表現パターン

- 文型パターンの記述 (864パターン)  
- CLV^renyou なさいっ。
- 抽象的概念  
- 聞き手に行動を要求する
- 応用のための情報  
- 話し手の情緒成分  
喜び(0.15) 悲しみ(0.3) 怒り(0.44)  
- 要求する行動の内容 = CLV

16

## 「鳥バンク」の場合 (池原2004)

- 日本語から英語への翻訳に特化
- 抽象的概念は特定せず、線形要素を特定してパターン化
- 重文複文の抽象的概念のラベル化が要改良 (現在は構文よりのラベル)
- 線形要素 = 単語レベル, 句レベル, 節レベル (23万パターン対)

日文: 明け方 になって やっととろとろ眠った。

英文: I dropped into a fitful doze at dawn .

日パ: TIME1 になって <N1は> やっととろとろ眠った。

英パ: N1 dropped into a fitful doze at TIME1 .

17

## 具体例

```

UJ000020-00:大阪に着くまでの間、ずっと本を読み続けた。
LE000020-00:read until I reached Osaka.
WJ000020-00:/y</tkN1(NI:5)は>/tcfkN2(OR:大阪,NI:405,NI:414,NI:415,NI:416,NI:464,NK:10,NK:17,
NK:27,NK:65,NK:98)|に/cfV3(OR:NY:1801,NY:2301,NY:2304,IY:8420)^rentai(まで|迄)の間、
</N1(NI:5)[ま>#5[「ずっと」]
IY:2510)|に
WE000020-00:
PJ000012-00:/の!間、</N
IY:2510)|に
PE000012-00:
CJ000004-00:/の!間、</N1(NI:5)は>/tcfkN2(OR:大阪,NI:405,NI:414,NI:415,NI:416,NI:464,NK:10,NK:17,
NK:27,NK:65,NK:98)|に/cfV3(OR:NY:1801,NY:2301,NY:2304,IY:8420)^rentai(まで|迄)の間、
CE000004-00:
1 副
AD2068 副
CL2 【<従属接続

```

18

## 具体例 - 意味的制約カット

LJ000020-00:大阪に着くまでの間、ずっと本を読み続けた。  
 LE000020-00:I read until I reached Osaka.  
 WJ000020-00:/y</tkN1[は>/tcfkN2[に/cfv3^rentai(まで|迄)の間、  
 <!N4[は>#5[!ずっと]/tck本を/cfv6(つづけ|続けた)。  
 WE000020-00:<I|N4> V6^past until <I|N1> V3^past N2.  
 PJ000012-00:/y</tkN1[は>|VP2^rentai(まで|迄)の間、<!N3[は>#4[!  
 ずっと]/tck本を/cfv5(つづけ|続けた)。  
 PE000012-00:<I|N3> V5^past until <I|N1> VP2^past.  
 CJ000004-00:/y/tcfkCL2^rentai(まで|迄)の間、!CL4.kako。  
 CE000004-00:CL4^past until CL2^past.  
 1 - 副 8420 - 2121,2220 FUa203 - まで  
 の間-  
 A02068 従属接続詞相当 until/till  
 CL2 [<従属接続詞相当> CL1]. CL2 [(until|till) CL1].

19

## 結果

- 確認できたこと
  - 文型パターンの可能性
    - 言語表現に表される抽象的概念が捕獲可能
    - パターンに付随する応用の情報で意味処理に繋がる
  - 文型パターンの記述言語仕様
    - 字面, 変数, 関数, 記号
    - 日本語表現の解析可能
    - 英語表現の生成可能
- 未解決
  - より効率的なパターン知識の作成
    - 判断基準の明確化
    - 作成の自動化

20

## 5. デモ

- 重文複文文型パターン辞書(鳥バンク)を用いた日英機械翻訳
- 辞書の規模
  - 単語レベルパターン: 121,913
  - 句レベルパターン: 79,476
  - 節レベルパターン: 25,482
- 適合したパターンを手動で選択して英文生成

21

## 参考文献

- 三浦つとむ: 日本語はどういう言語か, 講談社, 1976.
- 池原悟: 自然言語処理の基本問題への挑戦, 人工知能学会誌, 16(3), pp.422-430, 2001.
- 池原悟: 非線形言語モデルによる自然言語処理, 基礎と応用, 岩波書店, 2009.
- 田中努, 徳久雅人, 村上仁一, 池原悟: 結合価パターンへの情緒生起情報の付与, 言語処理学会第10回年次大会, 発表論文集, pp.345-348, 2004.
- 徳久雅人, 前田浩佑, 村上仁一, 池原悟: 対話行為と情緒を解析するための文末表現パターンの作成, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション, NLC2007-95, pp.45-50, 2008.

23