

## 用例に基づく換言 中日旅行会話翻訳への適用

大竹 清敬  
kiyonori.ohatake@atr.co.jp.

ATR 音声言語コミュニケーション研究所  
2003年3月19日

- 中日音声翻訳
- 換言に注力した翻訳方式：  
翻訳の前後に換言を適用  
→ 頑健性の向上
- 換言 → 単言語処理：  
言語対に非依存  
資源が豊富

用例に基づく換言 - p.176

用例に基づく換言 - p.176

## 目的

より自然な表現へ換言

例)

部屋を予約するをしたい  
↓  
部屋の予約をしたい

用例に基づく換言 - p.176

用例に基づく換言 - p.176

## 音声翻訳における換言手法

- 規則に基づく方法  
頑健性の維持が高価
- 統計的な方法 (SMT)  
速度面で不安、制御しにくい
- 用例に基づく方法  
制御が容易、用例の収集は容易

用例に基づく換言 - p.176

用例に基づく換言 - p.176

## 用例に基づく換言

基本的には用例に基づく翻訳と同じ

**収集** 目標とする表現を収集

**検索** 被換言表現に類似した用例を探す

**適用** 被換言表現に用例を適用

用例に基づく換言 - p.176

## 収集

目標とする表現を収集

- 汎化：固有名詞などの不一致を回避  
ATRの大竹です → PのPです  
3月19日水曜日の予約ですね  
→ MN日Wの予約ですね
- 換言：用言の不一致を回避  
満室となっております → 満室となっています  
満室です

用例に基づく換言 - p.176

用例に基づく換言 - p.176

## 検索

1. 汎化・検索単位の抽出
2. 検索・スコア計算  
検索単位が多くの文に出現 ⇒ 重要度低い  
文の長さが大きく異なる
3. 編集距離計算  
検索単位の順序を考慮
4. 評価

用例に基づく換言 - p.176

## 適用

選択用例：汎化文字列を含む  
↓  
対応する文字列へ置換

例)

入力： ええ、鈴木様本日から三泊のご予約でございますね  
選択用例： はい、P様本日からN泊のご予約ですね  
出力： 出力： はい、鈴木様本日から三泊のご予約ですね

用例に基づく換言 - p.176

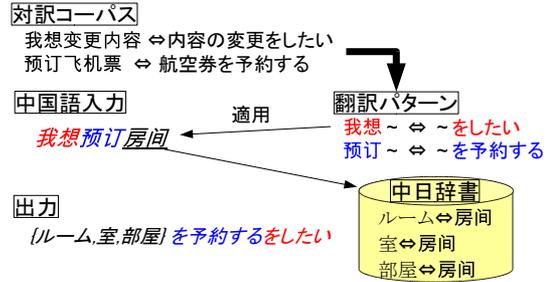
# 中日翻訳機

分類：EBMT

方法：対訳コーパスから翻訳パターンを抽出、翻訳パターンを適用し翻訳する

必要資源：中日対訳コーパス  
日本語形態素解析器  
中日辞書

# 中日翻訳機の概要



## 実験

### ● 中日翻訳機

学習文評価：2958 対訳

### ● コーパス

ATR 音声データベース+ATR 旅行会話基本表現集  
182000 件 → 汎化・換言により 337000 件

### ● 対象

翻訳結果と正解訳の編集距離が7以上の100発話

## 実験結果

合計	要改善	変化	改善	改悪
100	53	88	36	16

### 内訳

- (a) 名詞の置換が必要 (2)
- (b) 翻訳悪く、用例選択失敗 (6)
- (c) 表層的な違いにより用例選択失敗 (2)
- (d) 翻訳が不十分で用例選択失敗 (4)
- (e) 訳語選択失敗で用例選択失敗 (2)

## 換言事例

### 改善

ちょっと高いですねもっと安い部屋は持っていませんか  
→ ちょっと高いですね。もっと安い部屋は無いですか  
ええ仕事が終わったらお電話をお願いします  
→ はい、仕事が終わったらお電話します

### 改悪

バッテリーも売っていますか → バッテリーがあがっています  
きょうどんな様のショーがありますか  
→ きょうはどんな果物がありますか

## 考察

- 改善36 / 要改善53  
改悪例の多くは翻訳に起因
- 実験結果：楽観  
今後翻訳機に起因する誤りは減少

## 残されている問題

- 入力と選択用例の違い  
名詞の置換: シソーラスなどを用いる  
その他 (過不足部分など)
- 未知文評価 (オープンテスト)
- 大規模データベース構築時のシステムの挙動

## まとめ

- 換言に注力した中日音声翻訳システム
- 用例に基づく換言手法
- 小規模な学習文評価  
換言手法：楽観できる