

自然言語処理論I

14. 応用(機械翻訳とインターフェース)

1

機械翻訳

- machine translation (MT)
- ある言語の文を別の言語に自動的に翻訳すること
 - 原言語 (source language)
 - ◆ 翻訳する前の言語
 - 目標言語 (target language)
 - ◆ 翻訳した後の言語

2

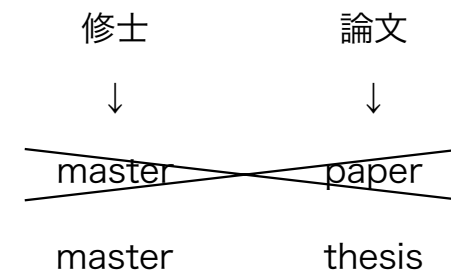
機械翻訳の主な問題点

- 2言語間のずれ
 - 語彙の対応のずれ
 - ◆ 英語: put on
 - ◆ 日本語: (服を)着る、(めがねを)かける、(帽子を)かぶる、(靴を)履く
 - 構造のずれ
 - ◆ 英語: SVO
 - ◆ 日本語: SOV
 - 明示的に表現する情報のずれ
 - ◆ 単数、複数の区別は日本語にはない
 - ◆ 単位の違い(一個、一台、一杯)は英語にはない

3

機械翻訳の主な問題点

- 非単調性
 - 文の部分訳を合成しても全体の訳になるとは限らない



4

機械翻訳の主な問題点

- 文脈依存性
 - 「論文」の訳し方
 - ◆ 論文 → paper
 - ◆ 修士論文 → master thesis
 - 上記のルールに当てはまらない場合
 - ◆ 彼はその大学で修士論文と博士論文の2つの論文を書いた
 - ◆ He wrote two theses in the university; the master thesis and the doctoral thesis.
 - 正確に翻訳するためには文脈に関する知識が必要

5

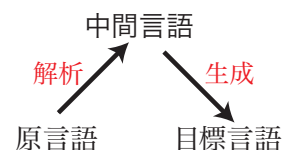
機械翻訳の方式

- 中間言語方式
- トランスファー方式 (変換方式)
- 実例型機械翻訳方式

6

中間言語方式

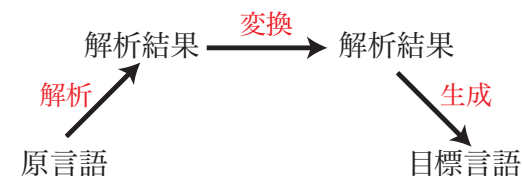
- 中間言語 (interlingua)
 - 文の内容の表現形式
 - 全ての言語について共通
- 処理の流れ
 - 解析 (原言語→中間言語)
 - 生成 (中間言語→目標言語)
- 特徴
 - 多言語翻訳への対応が容易
 - ◆ 言語と中間言語の相互変換さえできればよい
 - 中間言語の設計が非常に難しい



7

トランスファー方式

- 処理の流れ
 - 解析
 - ◆ 原言語の構文解析、意味解析
 - 変換
 - ◆ 解析結果を目標言語のものに変換する
 - 生成
 - ◆ 目標言語の文字列を生成する



8

トランスファー方式

● 特徴

- 「変換」は言語に依存
 - ◆ 個々の言語の組について変換規則を準備しなければならない
- 現実的なアプローチ
- 多くの機械翻訳システムで採用されている

● トランスファー方式のMTシステムの例

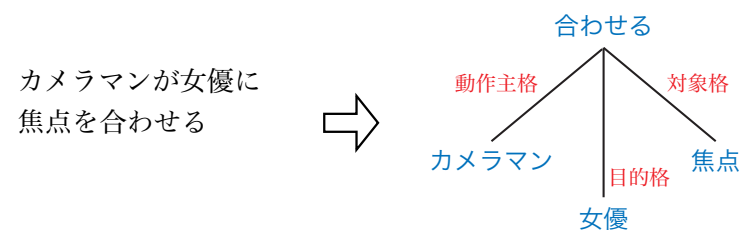
- Mu J/E, Mu E/J
 - ◆ Muプロジェクトによる日英機械翻訳システム
 - ◆ 科学技術論文の抄録の翻訳が目的

9

Mu J/Eの翻訳例

● 解析

- 原言語の文を格構造に変換する
 - ◆ 形態素解析
 - ◆ 構文解析
 - ◆ 意味解析 (格解析)



10

Mu J/Eの翻訳例

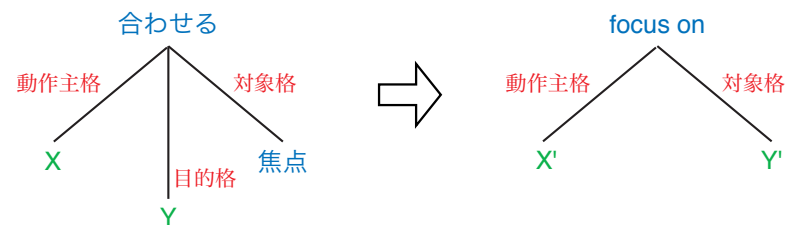
● 変換

- 原言語の格構造を目標言語の格構造に変換する
- 構造変換
 - ◆ 原言語と目標言語で格構造が異なる場合、適切な格構造に変換する
- 語彙変換 (訳語選択)
 - ◆ 原言語の単語を目標言語の単語に変換
 - ◆ 適切な訳語を選択する必要がある

11

Mu J/Eの翻訳例

● 構造変換の例



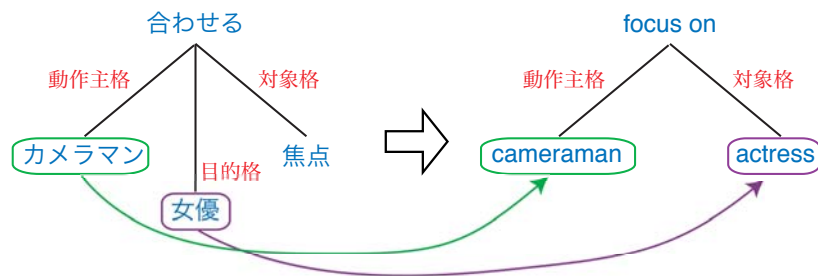
12

Mu J/Eの翻訳例

● 語彙変換の例

■ 「カメラマン」の訳語

- ◆ ○ cameraman (映画)
- ◆ × photographer (写真)



13

Mu J/Eの翻訳例

● 生成

- 格構造から目標言語の文を生成
- 主動詞の決定
- 深層格から表層格を決定
 - ◆ 英語の場合、名詞、動詞の語順が決まる



14

Mu J/Eの翻訳例

● 生成(つづき)

- 後処理を行う
- 冠詞、時制、数、人称の決定
 - ◆ The cameraman focused on an actress.
- 読みやすい英文への変換
 - ◆ 主語が長い場合はIt that構文に変換
 - ◆ 重文における重複要素の省略
Mary got up and ~~Mary~~ ate breakfast.

15

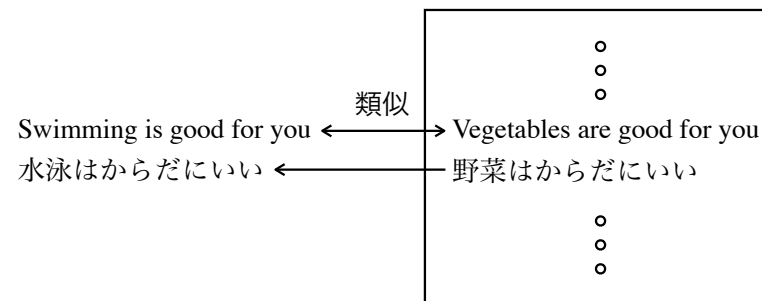
実例型機械翻訳

● 基本原理

- ある文と似たような文は同じように訳される
- 翻訳例を模倣して、翻訳文を生成する

● 例

翻訳例(パラレルコーパス)



16

実例型機械翻訳

● 処理の流れ

■ 翻訳メモリ(Translation Memory)を用意

◆ (大量の)翻訳事例の集合

◆ 翻訳パターンも同時に記述

ex. Vegetables are good for you ↔ 野菜は体によい

A be good for you ↔ A は体によい

■ 入力文と翻訳事例との類似度を計算し、もっとも似ている事例を見つける

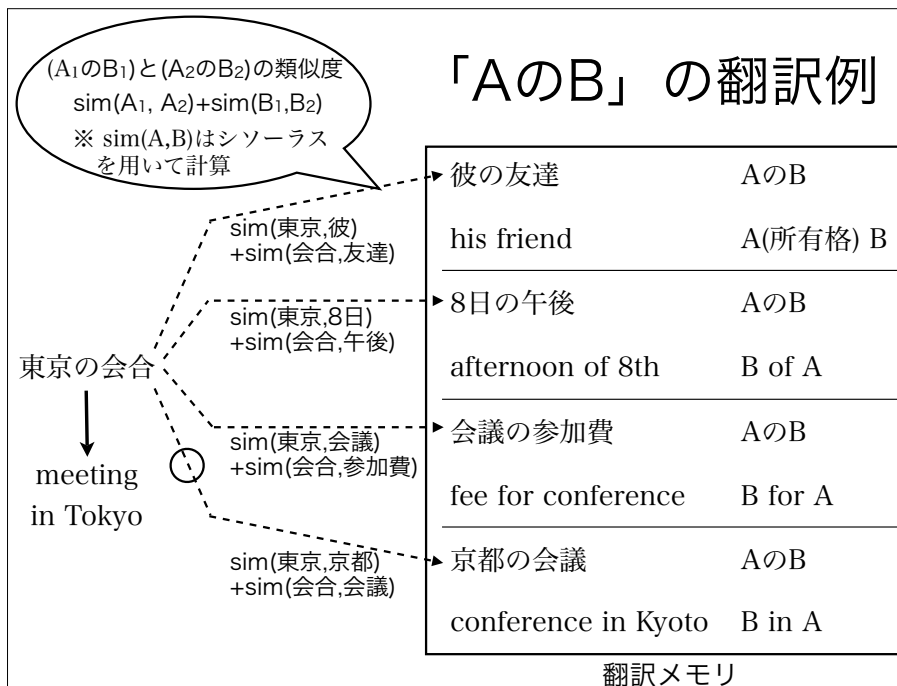
◆ 類似度はシソーラスなどを利用して計算

■ 翻訳パターンにしたがって入力文を翻訳

◆ 翻訳例の模倣

17

「AのB」の翻訳例



自然言語インタフェース

● コンピュータとのやりとりを自然言語で行うシステム

■ データベース検索のインタフェース

■ 対話システム (dialogue system)

◆ ホテルの予約

◆ 飛行機のチケットの予約

◆ 道案内システム

● 対話システムの入力

■ キーボード入力

■ 音声入力

19

対話システムの例

● 鶴来の飲食店案内

U: 鶴来にあるおいしいおそば屋さんを教えて

S: 草庵というお店があります

U: 場所はどこですか

S: 鶴来駅からの道順を案内すればよろしいでしょうか

U: はい

S: まず、駅を出て右に曲がって下さい

U: はい

S: 次に...

(Uはユーザ、Sはシステム)

20

対話システム

● 要素技術

- 音声認識
- 発話の理解
 - ◆ 形態素解析、構文解析、意味解析
- プランニング
 - ◆ 対話の内容を決める
- 意図理解
 - ◆ 相手の意図を理解し、適切な応答をする
- 談話構造の認識
 - ◆ 対話の流れを把握する

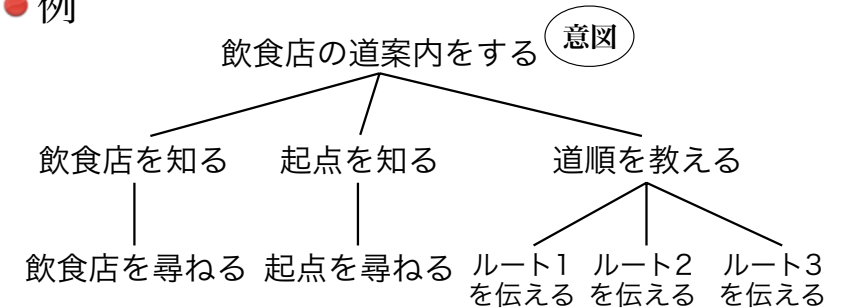
21

プランニング

● プラン (plan)

- 人間の行動: 欲求 → 意図 → 行動
- 意図の達成のために必要な行動計画

● 例



22

プランニング

● プランニング (planning)

- 意図をゴールとし、それを達成するプランを構築

● プランニングのための知識

■ (例) 因果関係の集合

- ◆ 飲食店を知る, 起点を知る, 道順を教える ⇒ 道案内をする
- ◆ ルート1を教える, 確認する ⇒ ルート1を伝える

■ ドメインの限定

- ◆ 世の中の知識を網羅的に記述するのは困難
- ◆ 対象を絞る (ex. 鶴来の道案内, ホテルの予約)

● 推論によるプランニング

23

意図理解

● 話し手の意図を理解する必要性

U: 島津先生の電話番号を教えてください
S: 島津先生は現在出張で不在です
✕ 電話番号はXXXXです

- 相手の意図(先生と連絡を取る)を正しく認識しなければ適切な応答はできない

● 計画認識 (plan recognition)

- 発話から相手の意図、行動計画を推測する
- プランニングと逆の操作
- 知識ベースを基に推論

24

談話構造の認識

- 一般的な対話
 - トピックが推移する
 - 同じトピックを持つ発話のまとまりが構造を持つ
→ 談話構造 (discourse structure)

[S: 北國銀行の交差点を右に曲って下さい
[U: すみません、北國銀行というと、JAIST
シャトルバスのバス停の近くですね } 部分対話 ※
[S: はい
U: それで?
[S: 次に...

※ 対話の本筋から一時はずれて、別の話題について話す対話

25

談話構造の認識

- 話題の推移をどのように検出するか?
 - ドメイン知識
 - ◆ 計画認識に基づく
 - 手がかり語(cue word), 合図句(cue phrase)
 - ◆ 新たな話題への移行、元の話題への復帰などを示唆する語や句
 - ◆ ex. すみません, ところで, わかりました
 - 韻律情報
 - ◆ ピッチ、抑揚、声の大きさ

26

まとめ

- 機械翻訳
 - 中間言語方式
 - トランスファー方式
 - 実例型機械翻訳方式
- インタフェース
 - 対話システム
- 自然言語処理における応用システム
 - 様々な要素技術の統合

27