

言語解析論

講師 竹内孔一

本日の内容

- 機械翻訳
 - 機械翻訳の問題設定
 - 機械翻訳システムの実現
 - トランスファ方式
 - 確率モデルによる方式

現段階の翻訳

- 問題の設定
 - 機械翻訳
 - 表層に現れている語を忠実に直訳
 - きまりきった表現の変換
 - ただし分野・文脈は前もって限定
 - 統計翻訳(ここ10年)
 - 翻訳支援
 - 作成した表現が使われているか
 - 複合語, 固有表現, きまった言い方 → Web でのチェック
- 分野・文脈を扱うモデルは研究中
- 例) 大型類人猿 → big anthropoid ape ??
great ape

翻訳

- (機械)翻訳での中心課題
 - 訳語候補生成と選択
 - 単語, 句, 文, どういう単位をどういう言葉に変換すればよいか
 - どう対訳辞書を構築するか
 - 訳語の並べ替え
 - 訳語をどの順に並べればよいか
 - 目的言語の文法ルール

翻訳システム

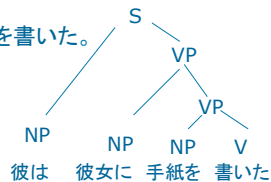
- 3つの変換方式
 - 中間言語方式 多言語化向き
 - 中間的な言語に一度変換してから目的語に変換する (意味記述の利用) 難しい
 - トランスファ方式
 - 構造を変換して目的言語に変換する 2言語間
 - 直接変換方式
 - 単語を直接変換
- 語順がほとんど同じ言語に適用 (例) 英仏, 日韓

トランスファ方式

- 前提
 - 翻訳する源言語と目的言語を決定しておく
- 特徴
 - よく利用されている方法
- 処理の流れ
 - 構文解析を行う
 - 構造の変換を行う
 - 目的言語に単語を置き換える
 - 目的言語での文法の修正を行う

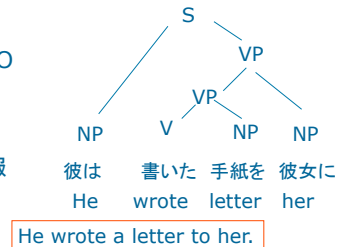
翻訳の例

- タスク
 - 日本語から英語を考える
- 入力文
 - 彼は彼女に手紙を書いた。
- 構文解析

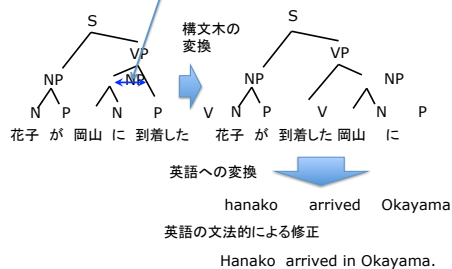


翻訳の例

- 構造の変換
 - SOV -> SVO
- 訳語の生成
- 動詞の情報
 - 動詞の格情報と英語の文法による修正



この部分を入れ替える



9

練習18

- 例文をトランスファー方式で英語に翻訳せよ

私は彼に手紙を渡した

I gave the letter to him.
(「手渡した」だと I handed the letter to him.)

私は彼の音楽に感動した

His music moved me.

こちらは変換が
難しい例

近年のやり方

- 統計翻訳モデル
 - 文単位で訳したparallel コーパスを作成
 - 統計的学習により変換を行う
 - P.F. Brown et al. (1993)

$$\hat{j} = \arg \max_j P(j|e)$$

$$= \arg \max_j \overbrace{P(e|j)P(j)}^{\text{翻訳モデル 言語モデル}}$$

11

統計翻訳モデルの特徴

- 簡略化
 - 単語の変換ではなく単語列単位
 - 手紙を出す/送る → send a/the letter
 - 文法ではなく単語列の確率 $P(j)$
 - もっともらしい単語の並び(5-gram, 10-gram)
 - 構文の誤りなどは考慮しない
- 評価
 - 正解コーパスとの差で評価
 - 翻訳の多様性は無視

12