



# 機械翻訳

---

**NHK放送技術研究所**

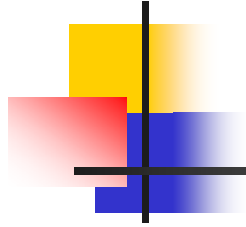
**加藤 直人**



# はじめに

---

- **私の経験と偏見に基づいて**
  - **NHKとATR('94.8-'97.7、'03.12-'06.3)**
  - **機械翻訳(規則翻訳、統計翻訳、用例翻訳)**
  - **音声認識の言語モデル**
- **講演概要**
  - **これまでの10年、これからの10年(?)**
  - **統計翻訳について**
  - **音声認識を参考にしながら**



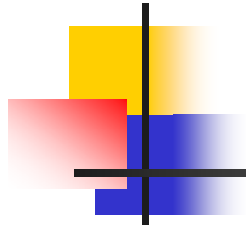
**これまでの10年**



# 機械翻訳のイノベーション

---

- この10年で機械翻訳の研究は大きく変化
  - 翻訳方式: 規則翻訳から**統計翻訳**へ  
(日本) (米国、欧州)
  - 翻訳評価: 主観評価から**自動評価**へ



# 翻訳方式



# 規則翻訳

---

- **規則に基づく翻訳(Rule-based MT)**
- **翻訳知識を手で作成**
  - 対訳辞書、解析規則、変換規則、生成規則...
  - 専門的な作業
  - 莫大な翻訳知識が必要
- **'80年代'、90年代に日本で盛んに研究された**
  - 現在は一段落
  - まだまだ不十分ではあるが...
- **PC上の商用翻訳ソフト、Webサイトで無料で**
- **問題は...**
  - **多大な開発コストがかかる**
  - **直訳調**



# 統計翻訳(1)

---

- **統計的機械翻訳(Statistical MT)**
- **翻訳知識を対訳コーパスから学習**
  - 機械翻訳の専門的知識は不要
  - 自然な翻訳になるかも
  - 対訳コーパスさえあれば、どんな言語対でも
- **'80年代末に出て、'90年代末から流行**
  - 翻訳需要(中国語→英語、アラビア語→英語)
  - 計算機性能の向上
  - ツールの公開(GIZA++, MOSES)

# 統計翻訳(2)

- P.F. Brown et al.

The mathematics of statistical machine translation: Parameter estimation.  
Computational Linguistics, 19(2) 263-311, 1993.

- Noisy channel **モデル**

$$\hat{e} = \arg \max_e P(e | f) = \arg \max_e P(f | e) P(e)$$

- 具体的には







# 統計翻訳(3)

---

- **現在ほぼ確立**

- **フレーズ(=単語列)単位の翻訳**

mes amis ⇔ my friends

vous présente ⇔ you introduce

vous présente mes amis ⇔ you introduce my friends

- **Log-linearモデル**

$$P(e|f) = \frac{1}{Z(f)} \exp \left[ \sum_m \lambda_m h_m(e, f) \right]$$

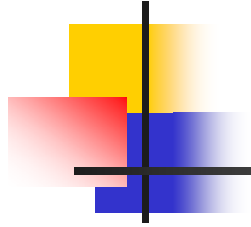
- **パラメータは評価指標BLEU(後述)に最適化**

- **言語モデルは5-gram、10-gram**

- **米国の機械翻訳コンテストが後押し**

- **DARPA, NIST**

- **精度はそろそろ頭打ち(?)**



# 翻訳評価



# 主観評価

---

- **翻訳の評価は難しい**
  - いろいろな表現がある
  - 人間によらなければならない
- **主観評価**
  - **適切さ(Adequacy): 原文の情報を含んでいるか？**
    - 5:All of the information, 4:Most of the information, 3:Much of the information, 2:Little information, 1:None of it (IWSLTより)
  - **流暢さ(fluency): 自然な表現か？**
    - 5:Perfect, 4:Good, 3:Non-native, 2:Disfluent, 1:Incomprehensible (IWSLTより)
  - **評価者によるばらつき**
  - **1段階評価を上げるのは大変！**



# 自動評価(1)

- 機械翻訳コンテストの評価のため
- 様々な評価方法を使う→言語の多様性を評価するため
  - BLEU (0-1.0)、NIST (0-∞)、WER (1.0-0)、PER (1.0-0)、...

- BLEUが良く使われる

- 正解訳とのn-gram (n=1,2,...,N)での一致度。通常N=4

$$BLEU = BP_{BLEU} \cdot \left( \prod_{n=1}^N \frac{s^*(n)}{s(n)} \right)^{\frac{1}{N}}$$

$BP_{BLEU}$  : 文が短い場合のペナルティ

$s(n)$  : 翻訳文中のn-gramの総数

$s^*(n)$  : 翻訳文中のn-gramで正解訳と一致した数

- n=小でAdequacyを評価、n=大でFluencyを評価
- わずかな差でも現れる→論文が書きやすい
- 規則翻訳に対しては低く判定するという批判も
- では、現状は……

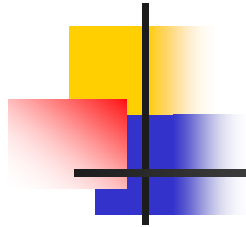


## 自動評価(2)

- NIST Open MT Evaluation
  - 単純には比較できないが、

### BLEU

	アラビア語→英語			中国語→英語		
'05年	0.513	0.466	0.465	0.353	0.307	0.300
'06年	0.503	0.459	0.449	0.363	0.297	0.285
'08年	0.456	0.453	0.451	0.309	0.306	0.304





**これからの10年**



# 未来予測

---

- そんなことわからない！！
- またもや 
  - 自動翻訳は人類長年の夢！
  - 研究は栄枯盛衰を繰り返す？
    - '50年代 スプートニクショック
    - '80年代 日本の国際化、海外への情報発信
    - '00年代 経済発展の中国、混迷を極める中東諸国
- それとも 
  - 音声認識の10年前と同じ(?)
    - 技術が確立され、ツールが充実
      - トライフォンHMM(音響モデル)、n-gram(言語モデル)
      - HTK、Julius...
    - 実用化へ
      - '00年 NHKのニュース音声認識
  - どうすれば...

## コーパスの規模？

	音響モデル $P(w y)$	翻訳モデル $P(e f)$
語数	40程度(音素)	数万~10万(単語)
単語列	数千程度(トライフォン)	???(フレーズ)
コーパス サイズ	数百万音素 (100時間)	NIST 中英 数億単語 (900万文)

数十億単語(数億文)!?





# 酔夢

---

- どうすれば...
  - 数億～数十億文の対訳コーパス
  - 計算機パワーのさらなる向上
  - もちろん手法そのものの発展も
    - 機械学習の応用として
- **統計翻訳の研究をやりましょう！！**
  - 統計翻訳の今後を知るには...
    - NLC研のシンポジウム(1/26-27 倉敷)
    - ユニバーサルコミュニケーションを実現するための言語処理技術
    - 招待講演「NICTにおける音声・言語研究拠点MASTARプロジェクトについて」