コーパスベースの機械翻訳

内山将夫@NICT
mutiyama@nict.go.jp

- コーパスとは単言語もしくは多言語のテキストの集積のこと
- コーパスベースの機械翻訳とは，コーパスから
- 各種の知識を自動獲得し，それを利用して，
- 機械翻訳をすること

背景

- コーパスベースの機械翻訳 (MT) は
- 学部実験で扱えるくらいに
- 各種のツールが整備されてきた

この講義の目的１

- 個人で機械翻訳を研究したい人が
- ツールのアルゴリズムを理解し，それを
- 改善し，新たなアルゴリズムを作成できるだけの
- 基礎知識を提供すること

この講義の目的２

- 自分でオープンソースの MT システムを利用して,
- 機械翻訳の実験をするくらい興味を喚起すること

成績評価方法

次の3通りのいずれかにより，成績を評価するので，各自が，どの評価法による評価を望むかを指定し，当該のレポートを提出して下さい．

評価法1 (MT実験) オープンソースのMTシステムを，各自の計算機で実際に動かしてみて，その過程をレポートする．

評価法2 (課題回答) 講義スライド中にある課題のうちで，興味のある課題について，レポートする．

評価法3 (授業態度) 講義の前に，講義スライドを全て読み，講義スライドにおける疑問点等を事前にリストアップし，講義においては，疑問点を質問し，そのやりとりの過程を記録する．そして，これらの疑問点とやりとりについてレポートする．

講義の内容

1. オープンソースの MT システムの例

---

2. MT の性能評価についての一般的な話題
3. MT の性能をどう測定するか？ 実験の作法と MT の自動評価
4. 自動評価尺度 BLEU

---

5. 初歩の確率
6. 初歩の言語モデル

---

7. 単語対応の導入
8. IBM Model-1 の式の説明
9. IBM Model-1 の動作例

---

10. 現状で利用可能なパラレルコーパス
11. パラレルコーパスを利用した検索と対数尤度比検定による対訳抽出
12. パラレルコーパスの自動作成

---

13. 翻訳モデルと翻訳エンジン
14. 句単位の翻訳モデルの概要
15. フレーズテーブルの作り方

---

16. 対数線型モデルの導入
17. 句に基づく統計的機械翻訳 (SMT) における素性

---

18. デコーダ概要
19. 最小誤り率訓練

---

20. まとめ