

## 20. まとめ

内山将夫@NICT  
mutiyama@nict.go.jp

# まとめ

まとめは課題とします

- 講義の内容をまとめたり
- Web 上で利用可能な機械翻訳システムの性能をレポートしたり
- その他，講義と関係のあることをレポートしたりして下さい。

## 今後の研究課題

- コーパス
- モデル/アルゴリズム/ソフトウェア
- 評価

## コーパス

対訳コーパスが必要だが，大規模な対訳コーパスは少ない

対訳コーパスを収集する効率的な方法が必要である。既に Web 等にあるものを収集することもできるが，その場合には，著作権を侵害しないようしないといけない。

対訳コーパスがない言語対については，最初から作らないといけない。そのときに，たとえば，標準的な日本語文を何文か用意しておいて，その文に対してさえ，対訳文を用意すれば，どんな言語対についても，コーパスベースの機械翻訳ができるようになれば，すばらしい。

## モデル/アルゴリズム/ソフトウェア

モデルは，現在は，何らかの形で構文を利用するものが主流である．けれども，単純なフレーズベースのものと比べて，特に性能が良いことが示されてはいない．単語対応をとるソフトウェアとして GIZA++ が広く使われているが，もうメンテナンスされていないし，性能も最高性能とはいえない．軽く速く動く単語対応をとるソフトウェアが欲しい．

## 評価

良い翻訳文とは何かが良くわかっていない。

翻訳文の評価は，人手によるものが最も信頼できると思われるが，人手による評価はコストも時間もかかる。簡単で，かつ，良い自動評価ができれば，機械翻訳の研究は，本質的に発展すると思われる。

## 結論

コーパスベースの機械翻訳は，まだ始まったばかりであり，これからもっと研究を進める必要がある．