

# AAMT長尾賞記念講演

6月17日

(独)情報通信研究機構多言語翻訳研究室

内山将夫・隅田英一郎

# お礼

NICTのSMTを評価していただいた方々  
研究・開発をサポートしていただいた方々に  
感謝いたします。

# SMT開発に必要なこと

- 対訳データ
- アルゴリズム
- コラボレーション

今回の受賞に関するSMTの開発の最初は10年以上前に遡るため、そのSMT開発に影響したと考えられることを順にご紹介します。

# 2000年には日英大規模対訳はなかった

- 英仏の大規模対訳データはあった
- フレーズベースSMTが研究された頃
- 2001-2003年度 科学技術振興調整費若手任期付研究員支援「言語横断文検索」を実施
- 同プログラムの一環として「日英新聞記事対応付けデータ (JENAAD)」を作成(18万文程度)
- 井佐原グループリーダー(現豊橋技科大教授)と読売新聞社に伺い、読売新聞と The Daily Yomiuriの研究利用を打診するとともに、JENAADの一般公開を許諾していただく。

# 2005年になっても大規模日英対訳はなし

- 2006-2010年度, 科学技術振興調整費 重要解決型研究等の推進, 「日中・中日言語処理技術の開発研究」(代表者: 井佐原均)
- 前半は日英・英日翻訳のEBMT  
300万文規模の日英対訳コーパスを作成したが一般公開はなし
- 後半は日中・中日翻訳のEBMT
- 全体を通じて、100万文規模の日中対訳コーパスの構築を進める
- The 1<sup>st</sup> Workshop on Asian Translation (WAT 2014) にデータが活用

# 日米特許対訳コーパスの作成(2007年頃)

- NTCIR-6において、藤井先生(現東京工業大学)らのご苦勞により、日米の特許10年間分が研究利用可能になる
- 常に対訳データの種を探していたので、早速、日米同時出願特許から対訳データを作成してみた → 180万文規模の対訳
- ようやく、英仏等と同等規模の対訳データができた。
- NTCIR-7,8の特許翻訳タスクを藤井先生、山本先生(筑波大学)、宇津呂先生(筑波大学)、内山で主催
- 特許翻訳において、規則ベースのMTの優位性が確認された

# 2010年ごろまでは、SMTでは英日・日英の長文翻訳精度は低かった

- NICTでは、旅行会話用の音声自動翻訳エンジンの開発を進めていて、2008年には実用化していたが、そのエンジンはフレーズベースを改良したもので、日英・英日の長文翻訳の精度は低かった
- NICTでは、2010年ごろに、アパレル等の特定分野向けのSMTエンジンを開発・実用化していたが、分野の特徴を利用した語順の並べ替えにより高精度化を図っていたため、一般的な長文の翻訳精度は低かった。
- NICTだけでなく、SMT全体でも、日英・英日翻訳における長文翻訳精度は低かった

# NTCIR-9,10で英日SMTの精度向上が確認

- NTCIR-9で、日本語が Head Final であることを利用して、前もって英語を語順変換し、それをフレーズベースで翻訳する方式の精度が、ルールベースの英日翻訳と匹敵か少し上回る性能を達成(2011年)
- NTCIR-10で、語順変換による英日SMTの手法が、ルールベースの手法の翻訳精度を明確に上回った(2012年)

**Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, Kevin Duh**

**Head Finalization: A Simple Reordering Rule for SOV Languages**

Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR, pages 244–251, Uppsala, Sweden, 15-16 July 2010



# 磯崎法にインスパイアされて語順変換を研究開発

- 任意言語対・任意言語方向に適用可能な自動学習方法を発明
- 翻訳性能が質的に向上(わからない→わかる)
- 以前の方法は、1000万文あっても汎用日英翻訳は無理と考えていたが、これ以降、対訳文があれば、ある程度の翻訳が可能になった

# パートナーとの共同開発

- 日本特許情報機構 (JAPIO)
  - 世界特許情報検索サービス
  - 従来中日MTエンジンと比較して3倍以上の翻訳性能
- 日本発明資料株式会社
  - MT PLUS
  - 要約1件あたりに含まれる誤り個数が従来方式の約1／12
- 凸版印刷株式会社
  - サポートラ
  - 翻訳効率を大幅に向上させる翻訳支援システム
- 川村インターナショナル
  - ITマニュアル専用の翻訳エンジン
- 株式会社バオバブ
  - 留学生ネットワーク@みんなの翻訳

# 明日公開：みんなの自動翻訳@TextTra<sup>®</sup>

- <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>
- NICT開発
  - 汎用日英・英日・日中・中日・日韓・韓日
  - 特許日英・英日
- 対訳文登録、用語登録によるカスタマイズ
- WebAPIの利用
- みんなの対訳をあつめてNICTのエンジンを再訓練

# 今後の展開

- SMTの技術は、頭打ちには程遠い状況です
- まずは、東京オリンピックに向けて
- グローバル・コミュニケーションの進展のためのパートナーをお待ちしています