

# 12

## あなた好みに翻訳する

～接頭辞制約によるニューラル翻訳の制御～



### どんな研究

ニューラル機械翻訳は、原文と訳文の関係を表す確率モデルを対訳データだけから学習するので、翻訳システムが出力する文をユーザが細かく制御することが難しいという問題があります。この展示では、ユーザがニューラル機械翻訳の出力を**カスタマイズする方法**を紹介します。

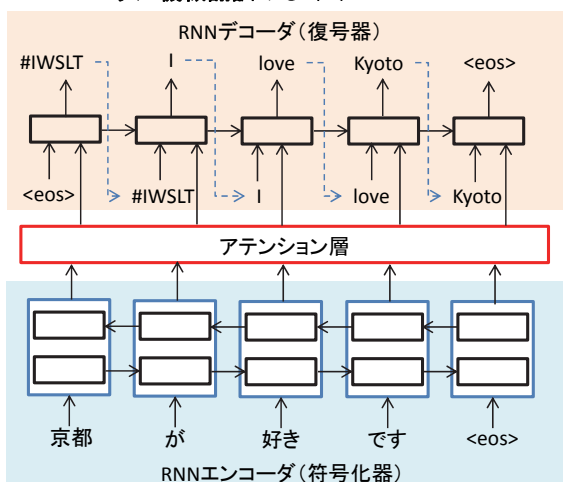
### どこが凄い

本研究では、翻訳システムが出力する文の特徴をタグで表現し、入力文から**タグと出力文を同時に予測**することで、タグで表現された特徴を反映した出力文を生成します。**ユーザがタグを指定**することによりそのタグの特徴を反映した出力文を生成することも可能です。

### めざす未来

翻訳の専門家は、文脈や状況、文化の違いなど様々な**言外の意味**を考慮して翻訳文を作成します。我々は、**文脈や状況を理解して**言外の意味を補いながら、最適な主語、構文、文体などを**自動的に選択**したり、ユーザへ適切な**選択肢を提示**するような機械翻訳システムを目指します。

#### ニューラル機械翻訳のしくみ



#### 接頭辞タグ(出力文の特徴)の例

京都が好きです → #3 I love Kyoto (文の長さ)  
 京都が好きです → #L2R I love Kyoto (左から右へ生成)  
 京都が好きです → #R2L Kyoto love I (右から左へ生成)  
 京都が好きです → #IWSLT I love Kyoto (旅行会話分野)

#### 日英翻訳の精度 (BLEU)

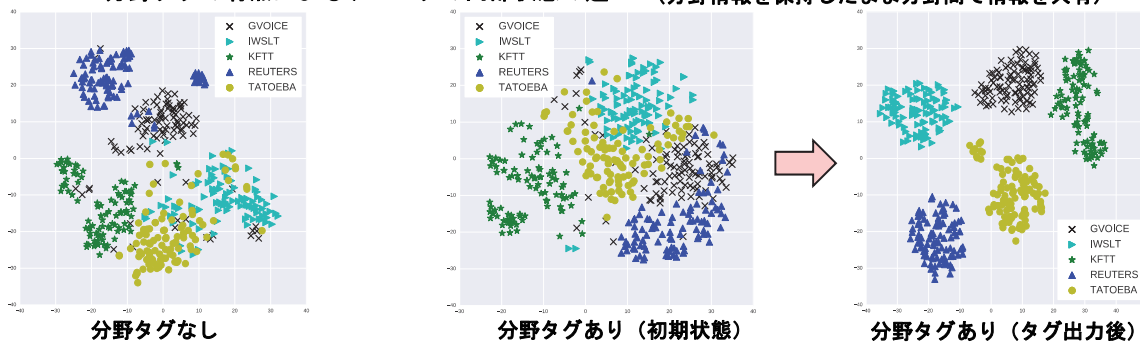
生成方向	IWSLT	KFTT	REUTERS
順 (#L2R)	34.8	20.9	19.7
逆 (#R2L)	32.8	20.1	19.6
方向予測	35.6	21.5	20.6

#### タグと出力文の同時予測の例

どういたしまして  
 → #IWSLT you're welcome  
 → #TATOEB A don't mention it

分野	単独	全体	分野予測	分野指定
GVOICES (ブログ, 4万文)	6.31	16.9	17.0	17.1
IWSLT (旅行会話, 2万文)	34.8	36.8	37.1	37.1
KFTT (Wikipedia, 43万文)	20.9	20.8	21.1	21.1
REUTERS (ニュース, 5万文)	19.7	24.6	25.0	25.0
TATOEB A (例文集, 19万文)	36.0	59.4	59.5	59.7

#### 分野タグの有無によるデコーダの内部状態の違い (分野情報を保持したまま分野間で情報を共有)



### 関連文献

- [1] 竹野峻輔, 永田昌明, 山本和英, “単語対応を利用した欠落語の投射による機械翻訳向きオラクル入力文の生成,” 電子情報通信学会技術報告, Vol.116, No.379, NLC2016-38, pp.135-140, 2016.
- [2] 竹野峻輔, 永田昌明, 山本和英, “ニューラル日英機械翻訳における欠落語に関する分析,” 言語処理学会第23回年次大会, pp. 659-662, 2017.

### 連絡先

永田 昌明 (Masaaki Nagata) 協創情報研究部 言語知能研究グループ  
 E-mail: nagata.masaaki(at)lab.ntt.co.jp