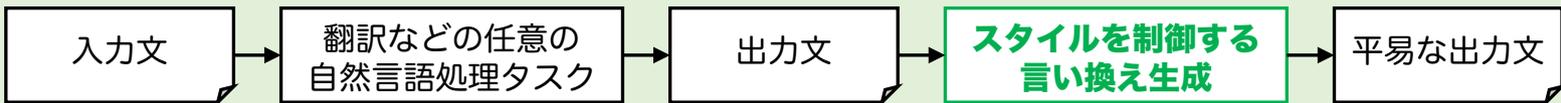


# 語彙制限に基づくパーソナライズされたテキスト生成

～ 好みに合わせたテキスト生成 ～ 梶原智之 (大阪大学) [kajiwara@jnlp.org](mailto:kajiwara@jnlp.org)



Please be careful -----> 厳重に注意してください -----> 気をつけてください

課題：自然言語処理において、出力文の意味的・文法的な正しさに加えて、**スタイルの良さ**も追求したい

解法：使用できる**単語を制限**することによってスタイルを制御できるテキスト生成モデルを開発する

成果：難しい→易しい・カジュアル→フォーマルのスタイル変換の性能を改善した (ACL-2019, AACL-2020)

## 出力単語を制御するスタイル変換 (ACL 2019), テキストの難易度制御 (ACL-SRW 2019, 言語処理学会論文誌)

$$PMI(w, s) = \log \frac{P(w, s)}{P(w)P(s)} = \log \frac{P(w|s)}{P(w)}$$

自己相互情報量：単語wとスタイルsの関連の強さを測る

Negative Lexically Constrained Decoding for Paraphrase Generation (ACL 2019)

$s \in \{\text{Complex, Simple}\}$        $s \in \{\text{Casual, Formal}\}$

- ① 自動 (PMI) または手動で難解な単語を検出
- ② 検出した難解な単語を使わずに言い換え生成

BLEU	Complex → Simple	Casual → Formal
先行研究の最高性能	24.3	74.5
RNN (Recurrent Neural Network)	24.1	74.7
RNN + 自動検出	<b>24.7</b>	<b>75.9</b>
RNN + 手動検出	36.4	79.3

Controllable Text Simplification with Lexical Constraint Loss (ACL-SRW 2019)

$s \in \{1, 2, 3, \dots, 12\}$

1~ 6: 小学1年生~小学6年生  
7~ 9: 中学1年生~中学3年生  
10~12: 高校1年生~高校3年生

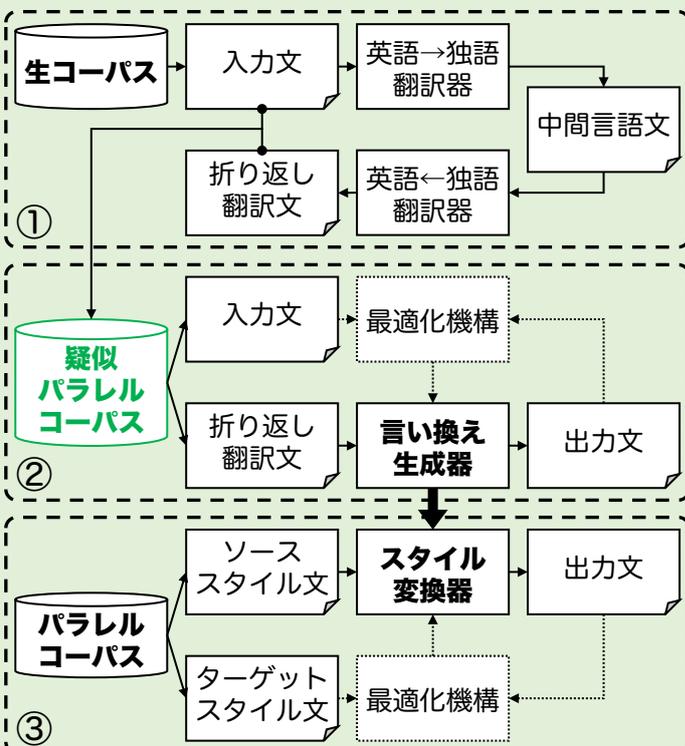
- (A) 文の難易度：文とともに目標学年を入力
- (B) 単語の難易度：CE損失をPMIで重み付け

	総合評価 (BLEU)	正解文と出力文のフレーズの一貫率	正解文と出力文の単語数の平均絶対誤差	出力文の単語と目標学年のPMIの平均値
RNN	20.4	4.38	0.12	
RNN + A	20.8	3.77	0.15	
RNN + AB	<b>21.9</b>	<b>3.69</b>	<b>0.19</b>	

高校生向け Verhasselt took the car to Houston last year for its first Shell competition, placing third of six hydrogen-powered vehicles.  
 中学生向け Verhasselt took the car to Houston last year for its first Shell **competition**.  
 小学生向け Verhasselt took the car to Houston last year for its first Shell **race**.  
 “competition”を指定 Verhasselt took the car to Houston last year for its first Shell **contest**.

論文は以下を参照してください  
[https://sites.google.com/site/mogura\\_nosenshi/projects/paraphrase-generation](https://sites.google.com/site/mogura_nosenshi/projects/paraphrase-generation)

## スタイル変換における少資源問題への対策 (AAAI 2020, 情報処理学会第241回自然言語処理研究会 若手奨励賞)



- ① 折返し翻訳：生コーパスから疑似パラレルコーパスを構築
- ② 事前訓練：汎用的な言い換え生成モデルを訓練 (スタイル非依存)
- ③ 再訓練：スタイルに特化した言い換え生成モデルに調整

入力文	折り返し翻訳文
I love <b>watching</b> the show.	I love <b>to see</b> the show.
Thanks for asking the question.	Thank you for the question.
The key to a successful relationship is <b>good communication</b> .	<b>Good communication is the key to a successful relationship.</b>

