

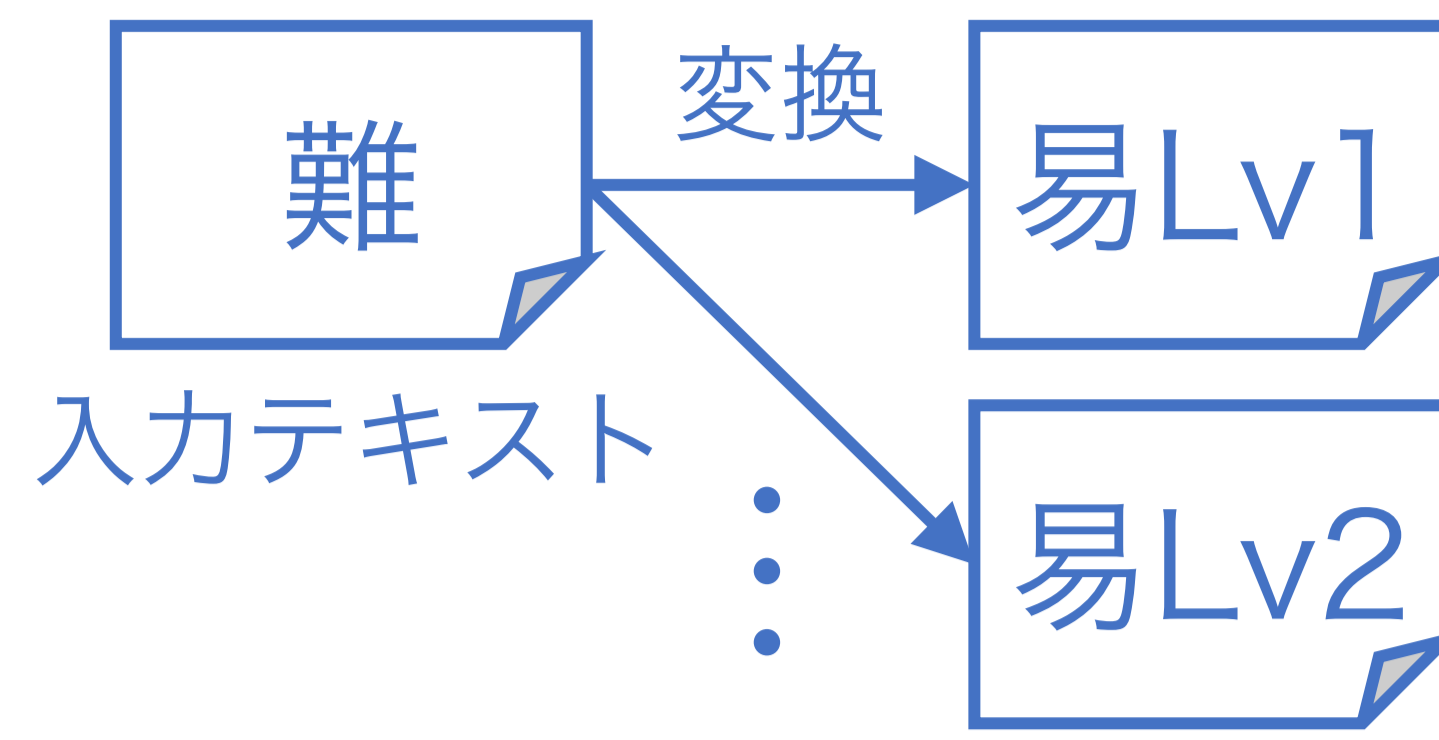
テキスト平易化における難易度の制御

西原大貴* 梶原智之† 荒瀬由紀‡

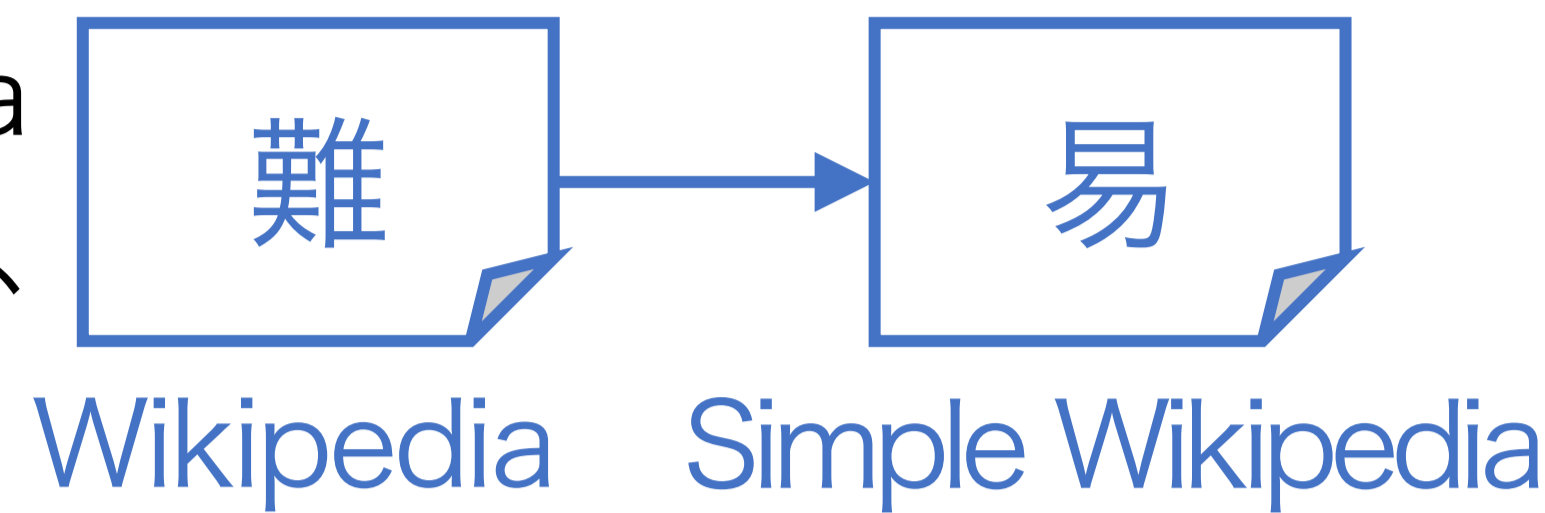
大阪大学工学部電子情報工学科* 大阪大学データビリティフロンティア機構† 大阪大学大学院情報科学研究科‡

背景・目的

言語学習支援のために、
テキストの難易度を学習者に
応じて変換したい



既存手法ではSimple Wikipedia
を用いて平易化の学習を行うが、
難易度のレベルは考慮しない



実験設定

- ◆ Marian-NMT (<https://marian-nmt.github.io/>)
- ◆ Newsela (各文書に2~12の難易度が付与)

	文書対数	うち文対数
Train	1,070	134,193
Valid	30	3,795
Test	30	3,594

システム概要



結果

	BLEU	SARI
Baseline-s2s	29.06	29.59
Multi-s2s	31.19	31.00

実例

成功例：元の意味を保持しつつ難易度に応じたテキストに変換された

入力	And as the teams led by giant weapons companies Northrop Grumman and Boeing Company try to win the support of lawmakers, the legislators want to know what their state will get out of the deal.
参照文	(7) Lawmakers mainly want to know what their state will get out of the deal. (4) They want to know what their state will get out of the deal.
Baseline	And as the teams led by giant weapons companies, the lawmakers want to know what their state will get out of the deal.
提案手法	(7) And as the teams led by giant weapons companies the support of lawmakers, the lawmakers want to know what their state will get out of the deal. (4) They want to know what their state will get out of the deal.

失敗例1：難易度に応じた変換はできているが、Leonardo da Vinciをpainterに変換するのは難しい

入力	He directs a museum dedicated to Leonardo da Vinci in his hometown of Vinci in Italy.
参照文	(5) He runs a museum dedicated to Leonardo da Vinci in his hometown of Vinci in Italy. (3) He runs a museum dedicated to the painter.
Baseline	He directs a museum dedicated to 600 million people in Italy.
提案手法	(5) He runs a museum dedicated to a museum dedicated to one of his hometown of Vinci in Italy. (3) He runs a museum dedicated to the museum.

失敗例2：難易度に応じた変換ができず、同じ文を出力してしまった

入力	She never had children, but later in her life received birthday cards from students all over the world, including a batch she kept from the Prairie View Intermediate School in Texas sent in 2000.
参照文	(9) Parks never had children, but later in her life received birthday cards from students all over the world. (7) Parks never had children, but later in her life she received birthday cards from students all over the world. (5) Later in her life, Rosa Parks received birthday cards from students all over the world.
Baseline	She never had children, but later in her life received birthday cards from students all over the world.
提案手法	(9) She never had children, but later in her life received birthday cards from students all over the world. (7) She never had children, but later in her life received birthday cards from students all over the world. (5) She never had children, but later in her life received birthday cards from students all over the world.

今後の課題

- ◆ 意味の等価性の保持
- ◆ 文の難易度推定
- ◆ デコーダに難易度情報を渡すなど他の手法との比較