4W-02

複数翻訳候補に基づく日本語若者言葉の対訳推定

新子 大晴* 譚 凱迅* 王 嘉慧♭ カリサ ティファニー* 李 亜森♭ 王 璐瑶♭ 延澤 志保*
*東京都市大学知識工学部 *東京都市大学情報工学部 ♭大連交通大学外国語学院

1 若者言葉の翻訳の現状

本研究は，日本語の若者言葉に対し適切な対訳語句を得ることを目的とする．

本稿では，各若者言葉を含むツイートをそれぞれ 50 ツイートずつ Twitter[1]から収集した．翻訳器は，複数の翻訳器について検証した上で，高い精度と知名度を有する [1] Google 翻訳[2]と DeepL 翻訳[3]の 2 種類を用いる．表 1 の中国語翻訳例では，翻訳器ごとに異なる訳語を選択している．図 1 の中国語での抽出語句を見ると，Google 翻訳では「emo」が，DeepL 翻訳では「情绪化」の割合が高いが，語句の種類はいずれもばらつきが見られる．英訳では，Google 翻訳と DeepL 翻訳どちらも「emo」が大半を占めており， 英語では中国語と異なり翻訳器による差が小さい (図 2)．これは言語の差とは言えず，若者言葉によって翻訳結果は異なる．「ズッ友」の中国語訳抽出語句について，Google 翻訳では「朋友」が 9 割以上を占めている一方，DeepL 翻訳では「朋友」と「zutto」が多いが他の語句も多数出現している (図 3)．英訳では DeepL 翻訳で「zutto」が最も多く 4 割を占める一方他の候補も多数出現しており，図 2 とは異なる (図 4)．

翻訳自体に失敗する文が多く訳語候補の推定に失敗する事例もある．「すこ」の中国語訳抽出語句の場合，Google 翻訳では「(不明)」が 7 割以上を占めている一方，DeepL 翻訳では 4 割未満であり，「喜欢」といった語句も出現している (図 5)．「すこ」の英訳では，図 5 と同様 Google 翻訳で「(不明)」が 7 割以上を占め，図 2 とは逆に DeepL 翻訳では候補が多数出現している (図 6)．

表 1: 「エモ」の中国語訳例

原文	エモめな写真を撮ってもらえた	抽出候補
Google	我拍了一张 emo 照片	emo
DeepL	他们给我拍了情绪化的照片。	情绪化

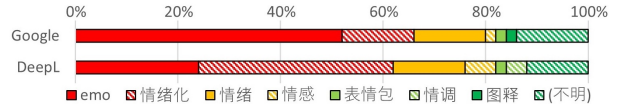

図 1: 「エモ」の中国語訳からの語句抽出 (翻訳器別)

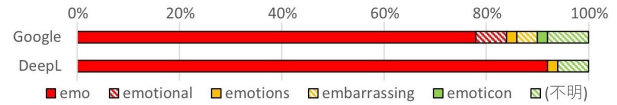

図 2: 「エモ」の英訳からの語句抽出 (翻訳器別)

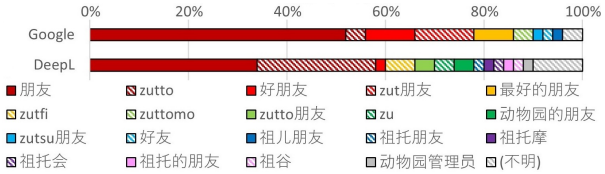

図 3: 「ズッ友」の中国語訳からの語句抽出 (翻訳器別)

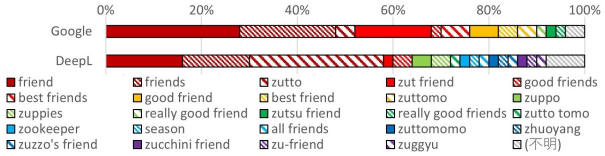

図 4: 「ズッ友」の英訳からの語句抽出 (翻訳器別)

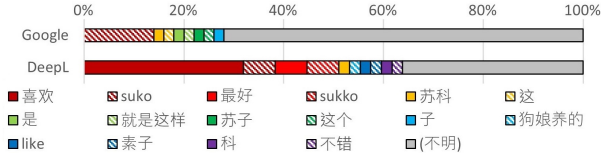

図 5: 「すこ」の中国語訳からの語句抽出 (翻訳器別)

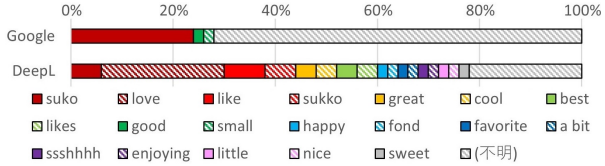

図 6: 「すこ」の英訳からの語句抽出 (翻訳器別)

このように，若者言葉の翻訳については，翻訳器ごとのその若者言葉の学習状況に応じて結果が異なる．

2 複数翻訳候補に基づく若者言葉の対訳推定

2.1 複数翻訳候補からの対訳推定

提案手法の基本的な方針は，翻訳対象とする若者言葉について，これを含む文を複数収集し，それぞれを複数の翻訳器に掛けることで得た十分な数の対訳文に含まれる共通要素を，この若者言葉の訳語と仮定するものである．複数の翻訳器による訳文間の類似度や，訳文の自然さなどに着目して対訳文を選択する手法 [2, 3] を基に，本研究では若者言葉から対訳語句への翻訳手法を提案する．提案手法では各若者言葉について複数の文を目的言語へ翻訳し，逆翻訳を利用して対訳語句を推定する (図 7)．

提案手法では，対象とする若者言葉の訳語が複数の対訳文中に高い頻度で出現することを仮定している．しかし，既存の翻訳器ごとに訳語が異なったり，翻訳に失敗したりする例が多い．そのため，出現頻度の高い訳語候

Japanese Teen Slang Translation from Multiple Candidates.
Taisei Atarashi*, KaiXun Tan*, JiaHui Wang♭, Tiffany Carissa*, YaSen Li♭, LuYao Wang (王 ロ瑶)♭, and Shiho Hoshi Nobesawa*.
* Faculty of Knowledge Engineering, Tokyo City University
* Faculty of Information Technology, Tokyo City University
♭ School of Foreign Languages, Dalian Jiaotong University

[1]Twitter, https://twitter.com.
[2]Google 翻訳, https://translate.google.co.jp.
[3]DeepL 翻訳, https://www.deepl.com/ja/translator.

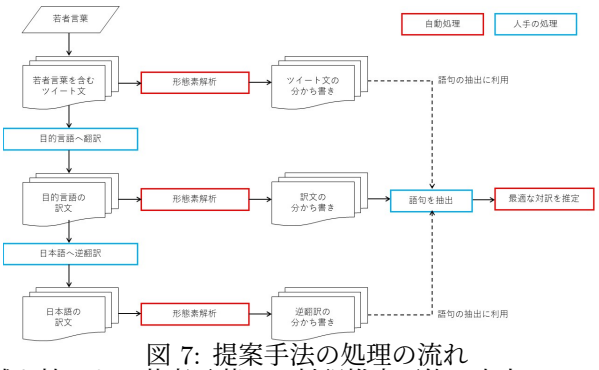

図 7: 提案手法の処理の流れ

補を持たない若者言葉は，対訳推定不能と出力することが妥当と考える．

2.2 若者言葉訳語選択実験

対象とする若者言葉は，若者言葉に関する記事を取扱う numan[1]で，2018 年から 2022 年までにランキングで挙げられたもののうち，Google トレンド[2]での過去 5 年間の検索出現頻度上位 10 個とした．このうち「エモ」「リセマラ」「ズッ友」「リアコ」は語を省略した短縮型 [4]，「推し」「ブーメラン」「マウント」「不可避」は本来の意味や用法とは異なる意味拡大型 [4] と考えられる．

訳語推定には，各若者言葉について，50 ツイートをそれぞれ 2 翻訳器で対象言語に翻訳して得た 100 文を用いた．表 2 に各翻訳器で最も多かった語句とその出現割合を示す． 提案手法では最も出現割合の高い語句を出力と

表 2: 若者言葉・目的言語ごとに推定した対訳語句

若者言葉	目的言語	各翻訳器最多候補 Google	頻度	DeepL	頻度	全体最多候補	頻度	システム出力	頻度
エモ	中	emo	52%	情绪化	38%	emo	38%	**emo**	52%
	英	emo	78%	emo	92%	emo	85%	**emo**	92%
リセマラ	中	重新滚动	38%	恢复	22%	重新滚动	19%	—	
	英	reroll	58%	reserch	18%	reroll	29%	**reroll**	58%
ズッ友	中	朋友	52%	朋友 zutto	各 34%	朋友	43%	**朋友**	52%
	英	friend	28%	zutto	28%	friend	22%	—	
リアコ	中	riako	32%	(不明)	30%	(不明)	26%	—	
	英	riako	52%	riako	16%	riako	32%	**riako**	52%
推し	中	(不明)	36%	猜测	54%	猜测	28%	**猜测**	54%
	英	(不明)	28%	guess	46%	guess	27%	—	
ブーメラン	中	回旋镖	84%	回旋镖	68%	回旋镖	76%	**回旋镖**	84%
	英	boomerang	80%	boomerang	82%	boomerang	81%	**boomerang**	82%
マウント	中	坐骑	54%	坐骑	26%	坐骑	40%	**坐骑**	54%
	英	mount	68%	mount	54%	mount	61%	**mount**	68%
不可避	中	不可避免	74%	不可避免	74%	不可避免	74%	**不可避免**	74%
	英	inevitable	56%	inevitable	70%	inevitable	63%	**inevitable**	70%
限界オタク	中	极限宅男	32%	边缘的怪人 边缘的书呆子	各 22%	极限宅男	16%	—	
	英	limit otaku	40%	marginal geek	36%	marginal otaku	20%	—	
すこ	中	(不明)	72%	(不明)	34%	(不明)	53%	—	
	英	(不明)	72%	love	24%	(不明)	47%	—	

する．その際，閾値を超える語句がない場合には対訳推定不能とし，表 2 では「—」と記述した．本稿では「(不明)」以外の各候補の出現割合の平均 (中国語 48%，英語 51%) を閾値とした．表 2 を見ると，多くの若者言葉で高い出現頻度を示す訳語候補が存在すること，また，これらの訳語候補は正解あるいは意図推定可能な訳語と見做せる可能性が高いことがわかる．これは事前に対訳候補を抽出することで不要な語句が除去された効果であると考える．複数の語句が現れる場合もあるが，文脈によって若者言葉のニュアンスが変化するため，むしろ適切な出力と考える．

[1]numan，https://numan.tokyo.
[2]Google トレンド，https://trends.google.co.jp/trends.

出力語句について，翻訳成功，意図推定可能，翻訳失敗の 3 段階で評価する．図 8 に訳語の評価を示す．縦軸

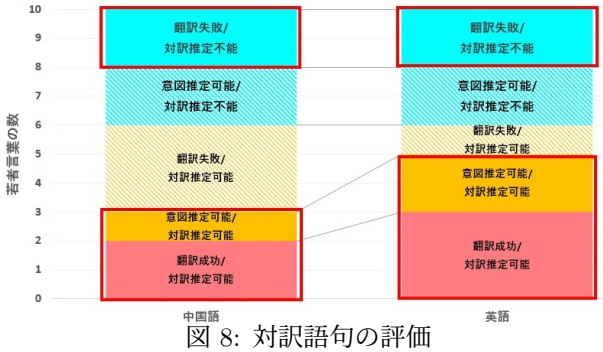

図 8: 対訳語句の評価

は若者言葉の数，横軸は翻訳した目的言語 (中国語と英語) を示す．ここでは，対訳を推定できた若者言葉を推定可能語句 (図 8 暖色)，できなかったものを推定不能語句 (図 8 寒色) と呼び，中国語と英語，ともに 6 割の若者言葉が推定可能語句だった (図 8 暖色)．推定可能語句のうち，赤枠で囲った部分について，中国語では，翻訳成功と意図推定可能を合わせた数が 3 個であり，翻訳の成功数を増やすための改善が期待できる．一方，英語では 5 個であり，意図推定可能の評価を成功として踏まえると，英語の方が翻訳器の正確性は高いと考える．

推定不能語句のうち赤枠で囲った部分は，翻訳結果に適切な訳語が含まれていない部分を指し，これについては対訳推定不能との出力が正解である．すなわち，提案手法が推定不能語句と判定したうちの半数は，既存の翻訳器での訳語推定が不可能であったことがわかる．

推定可能語句のうち翻訳成功と意図推定可能だった若者言葉，推定不能語句のうち翻訳失敗と出力された若者言葉，すなわち赤枠で囲われた全体の 60%に当たる部分がシステムとして適切な出力といえる．

3 まとめ

本研究では，留学生との円滑なコミュニケーションを目的として，翻訳の難しい日本語若者言葉の対訳を推定する手法を提案する．提案手法は若者言葉を含む文を入力として，複数の翻訳器で翻訳を行い，訳文から対となる語句候補を抽出し，その中から適切な対訳語句を推定するものである．この手法を用いて，中国語と英語を対象にそれぞれ翻訳し，日本語の若者言葉に対して適切な対訳の推定に成功した．

参考文献

[1] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,” arXiv:1609.08144v2, 2016.
[2] 山崎 亘涼, 孟 愛林, 張 文玉, 原田 千聖, 町田 翔, 延澤 志保, “日常会話を対象とした中日対訳文の自動選択,” 情報処理学会第 79 回全国大会, vol.2, pp.559–560, 2017.
[3] 中島 浩平, 何 婉瑩, 王 振章, 張 文玉, 町田 翔, 延澤 志保, “日常会話翻訳のための日中対訳文の自動推定,” 情報処理学会第 80 回全国大会, vol.2, pp.325–326, 2018.
[4] 菊池 つばさ, “浸透する若者言葉の特徴についての考察：由来となった単語との比較から,” 岩大語文, vol.17, pp.18–21, 2012.