

若者言葉を含む文の逆翻訳に関する一考察 A Study on Back-translation of Sentences Containing Teen Slang

真柄 圭佑*
Keisuke Magara

王 嘉慧*^b
JiaHui Wang

カリサ ティファニー*
Tiffany Carissa

延澤 志保*
Shiho Hoshi Nobesawa

1. 研究背景と研究目的

翻訳対象言語について十分な知識がない場合、翻訳結果の妥当性を推し量ることは困難である。そのため、逆翻訳結果を翻訳の妥当性の判断に用いることがある[1, 2]。逆翻訳後の文が原文と同様の内容を保持している場合、間の翻訳文も内容を保持することが期待される。

2. 研究に用いるコーパス

2.1. 日本語若者言葉コーパス

本稿では、新子による若者言葉翻訳の研究[3]を参考に、日本語若者言葉として表1の14語句を用いる。本

表 1: 実験対象とした日本語若者言葉

エモ	ズッ友	推し	不可避	すこ
リセマラ	リアコ	ブーメラン	マウント	限界オタク
蛙化現象	親ガチャ	草	えぐい	

稿では表1の各語句を含む文をTwitter¹から10文ずつ取得した合計140文を日本語若者言葉コーパスとする(図1)。表1の「エモ」は「エモい」という形容詞だが、

エモ	まさかの彼氏と七五三のスタジオ同じなエモすぎる
ズッ友	BABYMETAL とのズッ友写真。 イワチカたのむからズッ友でいてくれ

図 1: 日本語若者言葉コーパス中の文の例

名詞的に使われることも多く、複合語のような使われ方もある(図1)。また表1の「ズッ友(ずっと友達)」、「リアコ(リアルに恋してる)」など、独特な略語が多いことが日本語の若者言葉の特徴として挙げられる。

2.2. 中国語若者言葉コーパス

本稿では2022年に流行した中国語若者言葉²として表2の14語句を用いる。本稿では表2の各語句を含む

表 2: 実験対象とした中国語若者言葉

栓 Q	大冤种	小镇做题家	Pua	退! 退! 退!
团长	嘴替	一种很新的	刺客	电子榨菜
摸鱼	霸凌	女汉子	吃土	

文を微博³から10文ずつ取得した合計140文を中国語若者言葉コーパスとする(図2)。表2の「栓 Q」は英語の「thank you」の発音に漢字を当てた語で、直訳すると意味をなさない。また中国語の若者言葉には表2の「小镇做题家」のように言い回しがそのまま定着するものも多く[4]、翻訳が難しい(図2)。

3. 若者言葉の逆翻訳

本稿では、日本語と中国語それぞれの若者言葉について、逆翻訳結果を比較する。若者言葉コーパスに含まれる文それぞれに対して複数の翻訳器を用いて翻訳および逆翻訳を行い、特に逆翻訳で元の若者言葉が再現された例について考察する。

* 東京都市大学, Tokyo City University

^b 大連交通大学, Dalian Jiaotong University

¹Twitter, <https://twitter.com/>.

²The China Project, <https://thechinaproject.com/>.

³微博, <https://m.weibo.cn/>.

栓 Q	出门没带钱, 真的会栓 Q (お金を持たずに出掛けるとか, 本当自分に感謝しかないわ)
小镇做题家	每天都在加班, 栓 q 了 (毎日毎日残業だよ, ありがたいよまったく)
	小镇做题家们, 机会来了 (田舎のガリ勉のみなさん, チャンスだぞ)
	小镇做题家们只会做题不会出题 (田舎のガリ勉は問題解くばかりで問題作るのは無理)

図 2: 中国語若者言葉コーパス中の文の例

3.1. 日本語若者言葉の逆翻訳結果

若者言葉の逆翻訳結果について調査するため、日本語若者言葉コーパス(2.1節)について、既存の翻訳器を用いてそれぞれの文を中国語および英語に翻訳し、同じ翻訳器で日本語に逆翻訳を行った。本稿では、翻訳器としてBing翻訳⁴、DeepL翻訳⁵、Google翻訳⁶の3種類を用いた。

日本語若者言葉(表1)について、元の若者言葉が逆翻訳文に現れた割合を図3に示す。図3は、それぞれ

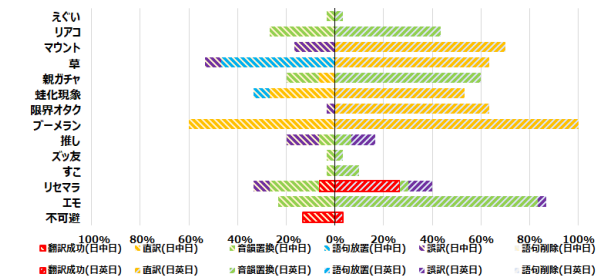


図 3: 日本語若者言葉の逆翻訳による再現

の若者言葉について、左に中国語を介した逆翻訳の結果、右に英語を介した逆翻訳の結果を示している。図3では、逆翻訳時に元の若者言葉が再現された理由を調べるため、中国語または英語への翻訳時の若者言葉の翻訳結果ごとに色分けしている。図3の赤は対象言語への翻訳時に正しい訳語に翻訳された割合、黄色は直訳となっていた割合、緑は元の文字列の音がローマ字などの形で表現されていたものの割合、水色は元の文字列がそのままの形で残った割合、紫は誤訳の割合を示す。例えば図3の「不可避」は、元の若者言葉が再現された割合は低い、そのすべてが正しく翻訳されたものである。それに対して、図3の「ブーメラン」は英語を介した場合100%、中国語の場合でも60%が元の若者言葉に逆翻訳されたが、これらはすべて、英語の場合は「boomerang」、中国語の場合は「回旋镖」に直訳された結果である。図3の「マウント」、「限界オタク」については、多くの文で英語では直訳の直訳として元の言葉に戻った。また、図3の「エモ」は、ほぼすべてが、中国語でも英語でもローマ字表記「emo」や「艾莫」などの音を当てた漢字表記に翻訳されたも

⁴Bing <https://www.bing.com/translator>.

⁵DeepL 翻訳, <https://www.deepl.com/translator>.

⁶Google 翻訳, <https://translate.google.com/>.

のが逆翻訳でカタカナに戻ったものであった。図 3 の「草」は、英語では「grass」と直訳されて逆翻訳で「草」に戻り、中国語では多くが中国語訳でも「草」のまま残りそれが逆翻訳でもそのまま出力されている。

このように、図 3 を見ると、逆翻訳で元の若者言葉に戻った場合でも中国語や英語への翻訳に成功したとは限らず、その理由は言葉によって明確に分かれる。

3.2. 中国語若者言葉の逆翻訳結果

3.1 節と同様に、中国語若者言葉コーパス (2.2 節) について、既存の翻訳器を用いてそれぞれの文を日本語および英語に翻訳し、同じ翻訳器で中国語に逆翻訳を行った。中国語の翻訳では、翻訳器として百度翻訳¹、DeepL 翻訳、Google 翻訳の 3 種類を用いた。

中国語若者言葉 (表 2) について、元の若者言葉が逆翻訳文に現れた割合を図 4 に示す。図 4 の左側は日本

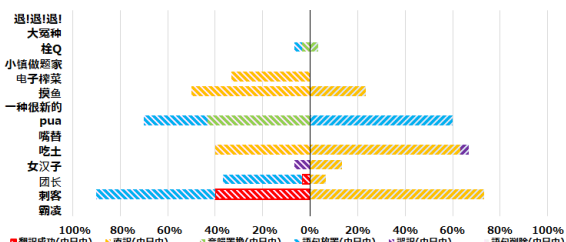


図 4: 中国語若者言葉の逆翻訳による再現

語を介した逆翻訳の結果、右側は英語を介した逆翻訳の結果を示す。図 4 も図 3 と同様の色付けをしている。図 4 の「团长」と「刺客」で日本語を介した逆翻訳の結果元の言葉が出力された割合が高いのは、これらの言葉が日本語でも同じ意味で用いられるためである。この場合、「团长」や「刺客」は翻訳としても正しい可能性があるが、図 4 の「pua」のように翻訳されずに残った可能性があること、また例えば中国語の「刺客」が日本語の「暗殺者」に翻訳され逆翻訳で「刺客」に戻った場合のように別の言葉に翻訳された場合と区別できなくなることから、赤や黄色ではなく水色で示している。

図 4 の結果から、日本語同様中国語でも、逆翻訳で元の若者言葉に戻った場合でも、それが日本語や英語への翻訳が成功したことを示すとはかぎらないことが示された。また、その理由は必ずしも日本語と同じではなく、言語によっても特徴があることが示唆された。

4. 若者言葉の翻訳

3 節では若者言葉部分の逆翻訳について、若者言葉部分が再現された状況について調査した。逆翻訳は元の他言語への翻訳結果を反映するため、ここでは逆翻訳前の翻訳文での若者言葉の翻訳状況について考察する。

日本語若者言葉コーパス (2.1 節) についての中国語、英語それぞれへの翻訳結果を図 5 に示す。図 3 と同様、図 5 の左側は中国語への翻訳結果、右側は英語への翻訳結果を示す。図 5 の白い部分は、翻訳の際に若者言葉に対応する部分が反映されなかったことを示す。日本語の若者言葉では翻訳の際に削除される場合が多く見られ、例えば「好き」の変形である「すこ」は中国

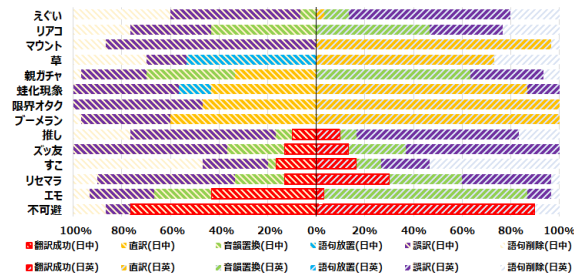


図 5: 日本語若者言葉の翻訳結果

語への翻訳でも英語への翻訳でもそれぞれ 53% が削除されている (図 5)。

中国語若者言葉コーパス (2.2 節) についての日本語、英語それぞれへの翻訳結果を図 6 に示す。図 4 と同様、

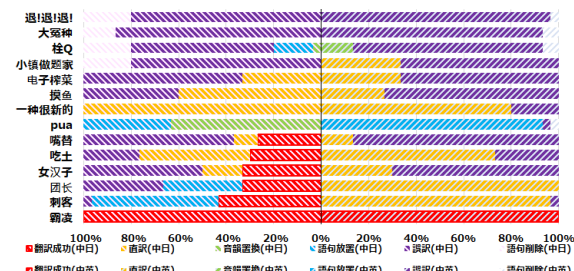


図 6: 中国語若者言葉の翻訳結果

図 6 の左側は日本語への翻訳結果、右側は英語への翻訳結果を示す。日本語若者言葉の翻訳結果 (図 5) と異なり、中国語若者言葉の翻訳では若者言葉がそのまま反映される例が多く見られた (図 6 水色)。中国語コーパスでは日本語コーパスと異なり白い部分がほとんどなく、代わりに誤訳が目立つ (図 6)。中国語コーパスの若者言葉は単語よりも句が多く [4]、翻訳不可能な状況が少ないことが一因と考えられる。

5. まとめ

本研究では、若者言葉の訳語の適切さの自動評価に向けて、日本語および中国語の若者言葉について原文と翻訳文、逆翻訳文の関係について調査した。その結果、逆翻訳文に元の若者言葉が再現された事例はほぼすべてが他言語への翻訳に失敗しており、意味を反映させた再現は一部の若者言葉のみでしか見られないことがわかった。この調査により、逆翻訳文と原文との比較は、若者言葉を含む場合については訳語の適切さ評価には有効でないことが明らかになった。

参考文献

- [1] 長瀬 友樹, 島津 美和子, 小谷 克則, 半田 峻一, 工藤 竜広, 葉文君, 中岩 浩巳, “実務翻訳における機械翻訳の利用に関する調査報告,” 言語処理学会 第 24 回年次大会, pp.742-745, 2018.
- [2] 弥永 啓子, “日本人大学生の機械翻訳使用の実態調査と今後の英語教育への導入に関する考察,” 京都橘大学研究紀要, vol.48, pp.1-19, 2022.
- [3] 新子 大晴, 譚 凱迅, 王 嘉慧, カリサ ティファニー, 李 亞森, 王 璐瑶, 延澤 志保, “複数翻訳候補に基づく日本語若者言葉の対訳推定,” 情報処理学会第 85 回全国大会, vol.2, no.4W-02, pp.815-816, 2023.
- [4] Tiffany Carissa, JiaHui Wang, Suravich Thammapattee, Keisuke Magara, Pasin Harisadee, Sivakorn Seinglek, Shiho Hoshi Nobesawa, “A Study on Teen Slang Translation of Asian Languages,” Forum on Information Technology (FIT2023), no.E-029, 2023.

¹百度翻译, <https://fanyi.baidu.com/>.