

日中翻訳のための若者言葉の対訳分析

何 婉瑋[‡] 藤田 和成* 浅沼 爽汰* 王 美琦[†] 郭 秀景[†] 中島 浩平* 町田 翔* 延澤 志保*

[‡]東京都市大学知識工学部 / 大連交通大学外国語学院 *東京都市大学知識工学部 [†]大連交通大学外国語学院

*東京都市大学大学院工学研究科

1 はじめに

近年、日本と中国の国際文化交流は、スマートフォンなどのSNSアプリを使って簡単にできるようになった。そのため、日中翻訳器の必要性が高まっている。翻訳器の精度は向上しているが、翻訳器による翻訳文は直訳や誤った翻訳が多く、特にSNSを最も多く使う若者たちが使っている若者言葉は正しく翻訳することは難しい。本稿では若者言葉を含む文を対象とした日中翻訳の対訳分析の結果をまとめる。

2 既存の翻訳器を用いた若者言葉の日中翻訳

米川は、若者言葉を「中学生から三十歳前後の若い男女が仲間内で娯楽・会話促進・連帯・イメージ伝達・隠蔽・緩衝・浄化などのために使う、規範からの自由と遊びを特徴に持つ特有の語や言い回し」と定義している [1].

複数の翻訳器の翻訳結果の比較と利用については、下里や山崎、張らの研究がある [2, 3, 4]. 本稿では3種類の翻訳器を用いて、若者言葉を含む文の翻訳の現状を調査した。分析対象は、最近の女子高生達がよくSNSで使う流行りの若者言葉ランキング77選 [5] から頻繁に用いられる若者言葉を10語(表1)選んでこれらを含む日本語文をTwitter[6]のツイート文からそれぞれ3文、計30文収集し、それぞれを3種類の翻訳器を用いて中国語に翻訳して得た中国語文90文を分析対象とする。Googleの検索エンジンから「日中 / 翻訳」で検索を行うと、Google翻訳、Excite翻訳、Weblio翻訳の3つの機械翻訳器が順に表示されたため、この3種類の翻訳器を用いた。

表 1: 分析に用いた若者言葉一覧

ワンチャン	リア充	テンアゲ	エンカ	チキる
イチキタ	キモい	ハズい	ガチで	とりま

この90文を4段階で評価した結果を図1に示す。ここでは文法も単語の対訳候補も適切なものを正解としたが、各翻訳器の出力には正解文は無かった。意味推定可能と

は、翻訳語句の選択や文法に誤りがあるものの元の文の意味を推定可能な文を示す。中国語の語句への翻訳に失敗しローマ字表記等を含んだものは翻訳失敗とし、それ以外の、翻訳語句の選択や文法の誤りが著しく元の文の意味の推定が困難なものを不正解とした。

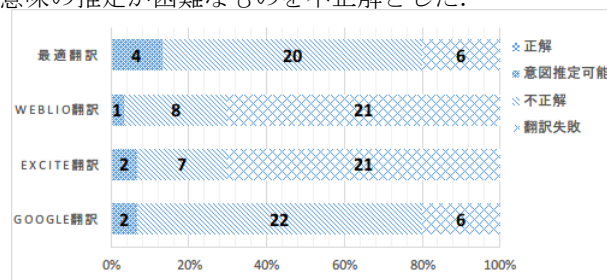


図 1: 複数の翻訳器の翻訳正解率

それぞれの日本語文に対し、3つの翻訳文のうち一番評価の高い翻訳をこの文の最適翻訳とした場合でも、日本語文30文に対して翻訳成功に当たる意図推定可能な文は4文に留まった(図1)。このことから、若者言葉を含む日本語文の翻訳が困難なことがわかる。

3 若者言葉の置換による翻訳精度向上

3.1 若者言葉から標準語への置換

本稿では、若者言葉の翻訳に際しての中国語句の選択が翻訳失敗の原因であると考え、翻訳の前処理として若者言葉を対応する標準語に置換することにより翻訳結果がどの程度向上するかを調査する。

翻訳に際して、若者言葉は品詞を考慮しながら対応する標準語へ置き換える。日本語俗語辞書 [7] には若者言葉を含めた俗語が2,418語、若者言葉辞典 [8] には1,020語収録されている。本稿ではこの2種類の辞書から若者言葉に対応する標準語を取得する(表2)。

表 2: 若者言葉の元の言葉と語義と置換した標準語

若者言葉	元の言葉	(品詞)置換した標準語
ワンチャン	ワンチャンス	(句)可能性は有る (名)可能性 (副)多分
リア充	リアルの生活が充実	(動)現実生活が充実している (名)現実生活が充実している人
テンアゲ	テンションアゲアゲ	(動)気持ちがあがる (名)興奮
エンカ	エンカウントする	(動)出会う
チキる	チキン	(動)おびえる
イチキタ	一時帰宅	(名)一時帰宅
キモい	気持ち悪い	(形)気持ち悪い
ハズい	恥ずかしい	(形)恥ずかしい
ガチで	ガチンコ	(副)本気で
とりま	とりあえず、まあ	(句)とりあえず、まあ

Analysis of Wakamono Kotoba for Japanese-Chinese Translation
WanYing He (何婉瑋)[‡], Kazumasa Fujita*, Sota Asanuma*, MeiQi Wang (王美琦)[†], XiuJing Guo[†], Kohei Nakajima*, Sho Machida*, and Shihou H.Nobesawa*.

[‡] Faculty of Knowledge Engineering, Tokyo City University / School of Foreign Languages, Dalian Jiaotong University

* Faculty of Knowledge Engineering, Tokyo City University

[†] School of Foreign Languages, Dalian Jiaotong University

* Graduate School of Engineering, Tokyo City University

久保村らは、若者言葉のうち省略型の若者言葉（例えば、「リア充」は「リアル充実」の省略）を辞書に登録されている語に変換する手法を提案している [9]。本稿では、標準語への置換の時点での失敗が無いよう、対応する標準語を手で確認している。

3.2 若者言葉から標準語への置換による翻訳精度向上

若者言葉から標準語への置換を行った結果、図2に示したように各翻訳器の正解率は明らかに向上し、最適翻訳

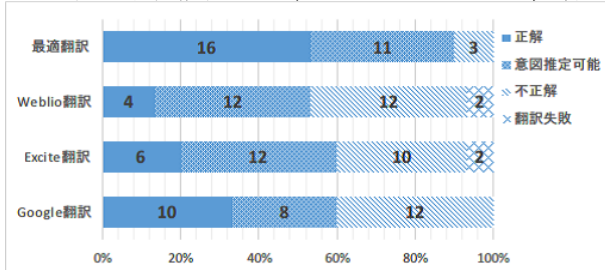


図2: 置き換え後の翻訳器の翻訳正解率

訳を選択した場合は30文中27文が翻訳成功、うち16文は正解となっただけでなく、最適翻訳での翻訳失敗が無くなった。図3に翻訳の具体的な例を示す。評価の「正解」「意図推定可能」「不正解」はそれぞれ○、△、×で表している。「ワンチャン」をそのまま翻訳対象とした場

例文	学歴悪くてもワンチャンあるけど青春は一度きりなんだぜ	評価	変換した文	学歴悪くてもワンチャンあるけど青春は一度きりなんだぜ	評価
Google	即使是不良的教育，也有一大块，但青春只有一次	×	教育背景即使不好，也有一次机会，但青春只有一次	即使是不良的教育，也有一次机会，但青春只有一次	○
Excite	学历即使坏也有一，不过青春一次只有啦	×	学历即使坏也有一机会，不过青春一次只有啦	学历即使坏也有一机会，不过青春一次只有啦	○
Weblio	虽然即使学历坏也有一张可是青春是一次	×	虽然即使学历坏也一机会可有可是青春是一次	虽然即使学历坏也一机会可有可是青春是一次	△

図3: 標準語への置換により翻訳に成功する例

合は3つの翻訳器にそれぞれ「一大塊 (大きな塊)」「一」「一张 (一枚)」に翻訳された。標準語の「ワンチャンス」へ置換してから翻訳した場合はそれぞれ「一次机会 (一回のチャンス)」「一机会 (ワンチャンス)」「一机会 (あるワンチャンス)」に正しく翻訳されており、翻訳成功となった。

3.3 若者言葉を含む文の翻訳の問題点

ここで、若者言葉の標準語への置換だけでは翻訳精度を向上させることができない例を検討する。

例文	死んだように眠るその顔はちょっとキモいが笑ってた	評価	変換した文	死んだように眠るその顔はちょっと気持ち悪いが笑ってた	評価
Google	我睡得像死的那张脸有点呆滞，但微笑	×	那睡得好像死的脸是有点恶心但是笑的	那睡得好像死的脸是有点恶心但是笑的	×
Excite	像死一样地睡的那个脸稍微令人讨厌，不过笑着	△	像死一样地睡的那个脸稍微难受，不过笑着	像死一样地睡的那个脸稍微难受，不过笑着	×
Weblio	像死了一样未开发的那个脸，有点恶心，但是笑了	×	像死了一样未开发的那个脸，有点恶心，但是笑了	像死了一样未开发的那个脸，有点恶心，但是笑了	×

図4: 標準語へ置換した文でも翻訳に失敗した例

1. 若者言葉が用いられている文の多くは話し言葉であり、文の構文要素 (主語、助詞など) は省略されることが珍しくない。主語などが略された文の中

国語に翻訳すると文の動作主が不明となり翻訳に失敗する場合がある。例えば、図4の「死んだように眠る」のは自分ではないが、「笑ってた」のは自分である。

2. 若者言葉を含む文の形態素解析に失敗することで、翻訳に失敗する場合がある。例えば、「チキる必要なんてないよ」という文は「必要ないよ」が「必要な」と「いよ」に誤分割され、「必要ある」とご認識されたことで原文と反対の意味に翻訳された。
3. 単語が多義語である場合、文意と一致しない意味に翻訳された。例えば、図4の「眠る」は「寝る」と「活用されていない状態」などの意味があり、ここでは誤った語義に対応する訳語が用いられたため翻訳に失敗している。図4では多義語は若者言葉では無いが、若者言葉も同じ問題を持つ。

若者言葉を含む文は話し言葉で記述されていることが多い。そのため、若者言葉をはじめとした辞書未登録語が多いこと、主語などの構文要素の省略が頻繁に起こることなど、翻訳は比較的難しい。それでも、若者言葉をそのまま翻訳対象とした場合 (図1) に比べて、若者言葉を標準語へ置換してから翻訳した場合 (図2) では明らかに精度が向上しており、若者言葉を置換することで翻訳が容易になることが示された。

4 まとめ

若者言葉を含む文について、機械翻訳器の翻訳精度はまだ低い。本稿では文中の若者言葉に対応する標準語に変換することによる翻訳精度の向上について調査し、若者言葉の置換が正解率の向上に寄与することを確認した。

参考文献

- [1] 米川 明彦, “若者ことば研究序説,” 大修館書店, 2006.
- [2] 下里 昌輝, 延澤 志保, “オノマトペを対象とした日英対訳語句の自動推定,” 電子情報通信学会 2016 年総合大会, vol.D-2, p.41, 2016.
- [3] 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.
- [4] 張 文玉, 町田 翔, 孟 愛林, 延澤 志保, “構文に着目した日中機械翻訳候補文の自動修正,” 情報処理学会研究報告, vol.2017-NL-232, No.7, 2017.
- [5] 最近の女子高生達がよく SNS で使う流行りの若者言葉ランキング 77 選【2017 年版】: <https://jikitourai.net/schoolgirl-use-expression>.
- [6] Twitter: <https://twitter.com/>.
- [7] 日本語俗語辞書: <http://zokugo-dict.com/>.
- [8] 若者言葉辞典～あなたはわかりますか?～: <http://boosesound.blog133.fc2.com/>.
- [9] 久保村 千明, 原田 俊信, 佐々木 輔, 山本 義人, 亀田 弘之, “ブログ記事を素材とする若者語処理システム評価方法,” 信学技報, Vol.105, No. 615, pp.165-169, 2006.