

機械翻訳のための日本語若者言葉の言い換えに関する一考察

何 婉瑋^{1,2,a)} ハダノウィチ アレクセイ¹ 藤田 和成¹ 浅沼 爽汰¹ 田村 亮介¹ 町田 翔³
山田 紘司³ 延澤 志保^{1,b)}

概要: 翻訳器の精度は向上しているが、若者言葉は正しく翻訳できるものがほぼない。日本語若者言葉を含む句を機械翻訳器に翻訳させる際、前処理として若者言葉部分を一般的な言い回しへ置換することで翻訳正解率は向上する。本稿では日本語の若者言葉を含む句を対象として言い換え処理を伴う機械翻訳を行った結果について、対象言語が中国語の場合とロシア語の場合それぞれ分析を行った結果をまとめる

キーワード: 機械翻訳, 言い換え処理, 日中翻訳, 日露翻訳.

A Study on Paraphrasing for Translation of Japanese Young People's Phrases

WAN YING HE^{1,2,a)} ALIAKSEI ALIAKSANDRAVICH KHADANOVICH¹ KAZUMASA FUJITA¹ SOTA ASANUMA¹
RYOSUKE TAMURA¹ SHO MACHIDA³ KOJI YAMADA³ SHIHO HOSHI NOBESAWA^{1,b)}

Abstract: Even though the accuracy of machine translation is getting better, it is not yet easy to translate young people's phrases correctly. For the translation of young people's phrases, paraphrasing into normal expressions is helpful in improving translation accuracy. We share the results of Japanese-Chinese translation and Japanese-Russian translation of these young people's phrases.

Keywords: Machine translation, paraphrasing, Japanese-Chinese translation, Japanese-Russian translation

1. はじめに

近年、スマートフォンなどで SNS アプリを使って簡単に国際文化交流ができるようになった。そのため、機械翻訳器の必要性が高まっているが、翻訳器の精度は大きく向上しているとはいえ、SNS を最も多く使う若者たちが使う口語表現については直訳など誤った翻訳が多いのが現状である。本稿では、日中翻訳と日露翻訳の 2 種類を対象として、若者言葉の翻訳について考察する。

2. 若者言葉から標準語表現への言い換え

2.1 分析対象とする若者言葉

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]。何らは若者言葉を含む文の翻訳に失敗が多いことに着目し、若者言葉の翻訳に際しての翻訳先言語の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に対応する標準語表現に言い換えることにより翻訳結果がどの程度向上するかを調査した [2]。

語彙の標準語表現への言い換えについては、語釈文の利用 [3] や俗語の処理 [4]、若者言葉のうち省略型の若者言葉を辞書に登録されている語に変換する手法 [5] などが提案

¹ 東京都市大学知識工学部情報科学科
Faculty of Knowledge Engineering, Tokyo City University

² 大連交通大学外国語学院
School of Foreign Languages, Dalian Jiaotong University

³ 東京都市大学大学院工学研究科情報工学専攻
Graduate School of Engineering, Tokyo City University

a) (何 婉瑋) enei17@ipl.cs.tcu.ac.jp

b) shiho@tcu.ac.jp

されている。何らの研究は若者言葉の日中翻訳を目的としており、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 [6] から頻繁に用いられる若者言葉を 10 語選んで、関連研究の手法を基に、若者言葉などの俗語を 2,419 語収録する日本語俗語辞書 [7]，若者言葉を 1,020 語収録する若者言葉辞典 [8] の 2 種類の辞書から若者言葉に対応する標準語表現を取得した (表 1)[2]。何らの研究で

表 1 省略型若者言葉とその言い換え表現 [2]

若者言葉	元の言葉	(品詞) 標準語表現
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) 気持ちが上がる (名) 興奮
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
エンカ	エンカウツする	(動) 出会う

は、標準語表現への言い換えの時点での失敗がないよう、若者言葉と標準語表現との対応を手で確認している [2]。

2.2 分析対象とする文

何らは、表 1 の 10 語を含む日本語文を Twitter[9] のツイート文からそれぞれ 3 文、計 30 文収集し、分析対象としている [2]。本研究でも、若者言葉を自然に用いている文としてツイート文が翻訳対象に適切と考え、何らの研究と同じ 30 文 (表 2) を分析対象とする。

ツイートは一般に話し言葉で記述されるため、表 2 にあるように、崩れた表現を多く含む。そのため、翻訳に際しては、言い換え対象とする若者言葉だけでなく、他の語句や構文にも問題がある可能性がある。本稿では、若者言葉の標準語表現への言い換えの翻訳への効果と、若者言葉を含む話し言葉文の翻訳の問題点について考察する。

3. 標準語表現への言い換えによる若者言葉の翻訳精度向上

3.1 複数の翻訳器による翻訳結果の比較

何らは、複数の翻訳器の翻訳結果の比較と利用についての研究 [10], [11], [12], [13] を基に、Google 翻訳, Excite 翻訳, Weblio 翻訳の 3 種類の翻訳器を用いて若者言葉を含む文の日中翻訳の現状を調査し、標準語表現への言い換えが翻訳精度の向上に有効であることを示した [2]。

日本政府観光局国籍/月別訪日外国人統計データ [14] に

表 2 分析対象とするツイート文

若者言葉	ツイート文
ワンチャン	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 もしかして、今日ちゃんと雪積もってくればワンチャン休みじゃ？ 部活行きたかったけどワンチャン風邪かもと思って先輩に迷惑かけるの嫌だから休んだ
リア充	意識してリア充になることは難しい 彼氏がいてリア充してるらしい リア充になる計画失敗だ
テンアゲ	最近お部屋めっちゃキレイだからテンアゲ こんなに沢山の雪見たの初めてだからすごいテンアゲだった 明日は初給料日でテンアゲやわー
チキる	チキる必要なんてないさ まあチキるよりいいでしょう めちやめちやチキるけど遊ぼう！
イチキタ	いちきたしてから行くわ イチキタしたしもう外出たくない 満足してイチキタしてしまった
キモい	死んだように眠るその顔はちょっとキモいが笑ってた 数年ぶりにキモい思いした キモい客しか来ない
ハズい	自分のパソコンのデータ見られるのとか超ハズいぜ 違っていたらハズイけど 何の話かわからんけどハズイ
ガチで	昨日はガチで疲れすぎてぶっ倒れてた これ年末ガチでやばい 最近腰がガチで悲鳴あげてる
とりま	とりま人多すぎるから会場から出る とりまこれからもよろしくお願ひします とりまこれ守ってれば痩せれるから
エンカ	またエンカした時は協力してね 駅で先生にエンカして気まずい 同じ建物にいてエンカしないって逆に凄い

よると、2017 年度に日本を訪れた外国人のうち総人数が一番多いのは中国人である。これに対して伸び率が一番高いのはメキシコ人、二番目はロシア人だが、ロシア人の総人数はメキシコ人より多いので、本稿では、翻訳先言語を中国語とロシア語の 2 言語に増やして標準語表現への言い換えの影響を考察する。

何ら [2] が用いた Weblio 翻訳器はロシア語が扱えないので、本稿では Google 翻訳, Excite 翻訳のほかにも中国語については中国語話者が主に用いる百度翻訳, ロシア語についてはロシア語話者が主に用いる Yandex 翻訳, それぞれ 3 種類の翻訳器による出力結果を比較検討する。

出力結果は表 3 の 4 段階で評価し、このうち正解および意図推定可能と判断されたものを翻訳成功、意図推定困難および語句放置を翻訳失敗とみなす。

表 3 翻訳結果の 4 段階評価

成	正解	語句も構文も適切なもの
功	意図推定可能	語句の選択や構文に不適切なものが含まれるが、原文の意図の推定が可能なもの。
失	意図推定困難	原文の意図の推定が困難もしくは不能なもの。
敗	語句放置	原言語の語句の翻訳に失敗しローマ字表記などの形でそのまま含んでいるもの。

日本語若者言葉 10 語に対応するツイート計 30 文 (表 2) と、この 30 文について若者言葉部分を対応する標準語表現 (表 1) に言い換えた 30 文とを対象としてそれぞれの翻訳器を用いて翻訳を行った結果を、図 1(日中翻訳)、図 2(日露翻訳) に表 3 の 4 段階にあてはまる文の数で示す。図 1

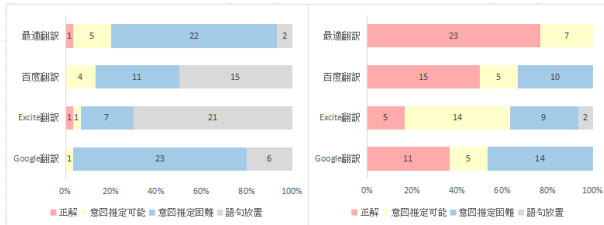


図 1 日中翻訳結果の変化 (左:若者言葉, 右:言い換え)

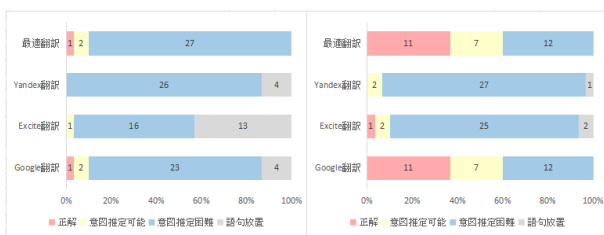


図 2 日露翻訳結果の変化 (左:若者言葉, 右:言い換え)

および図 2 のグラフのうち、左側は若者言葉を含むツイート 30 文 (表 2)、右側は若者言葉部分を標準語表現 (表 1) に言い換えた 30 文を示す。ここでは、複数の翻訳器の出力結果のうち一番評価の高い翻訳文をこの文の最適翻訳文とすることで各文についての言語ごとの比較を行う。

若者言葉をそのまま翻訳した場合、中国語への翻訳では日本語文 30 文のうち最適翻訳文として選ばれた文が翻訳成功となったものは 6 文で、うち正解文は 1 文に留まった。最適翻訳文でも 2 文は語句放置による失敗であった (図 1 左)。日露翻訳では、若者言葉をそのまま翻訳した場合最適翻訳文の中でも翻訳成功は 3 文しかなく、うち正解文も中国語の場合と同じく 1 文に留まった (図 2 左)。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。各翻訳器の出力結果に対する評価を比べてみると、日中翻訳の場合、翻訳成功の文は多いが、翻訳失敗となった文には語句放置で失敗した文も多い (図 1 左)。それに対して、日露翻訳の場合、翻訳成功の文は少ないが、語句放置による失敗の文は少なく、最適翻訳での語句放置はない (図 2 左)。

若者言葉から標準語表現への言い換えを行った結果、中国語の場合各翻訳器の正解率は明らかに向上し、最適翻訳を選択した場合には 23 文が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 1 右)。ロシア語の場合、若者言葉から標準語表現への言い換えを行っても各翻訳器の正解率は中国語の場合ほど向上していない (図 2 右)。最適翻訳を選択した場合には 30 文中 18 文が翻訳成

功、うち 11 文が正解となった。言い換え前と同じく語彙放置による翻訳失敗はなかった。なお、Google 翻訳は 2 言語双方で出力の評価がほぼ同様であった (図 1、図 2)。

3.2 標準語表現への言い換えによる翻訳成功の例

標準語表現へ言い換えを行った 30 文の翻訳結果を対象に、最適翻訳を選択した場合、日中翻訳と日露翻訳の双方で翻訳成功となった文は 18 文あった (表 4)。表 4 中、下線

表 4 標準語表現へ言い換えた結果翻訳に成功した文

- もしかして、今日ちゃんと雪積もってくれば多分休みじゃ？
意識して現実の生活が充実している人になることは難しい
- 彼氏がいて現実の生活が充実してるらしい
 - こんなに沢山の雪見たの初めてだからすごい気持ちがあがった
明日は初給料日で気持ちがあがるやわー
おびえる必要なんてないさ
まあおびえるよりいいでしょう
 - めちゃめちゃおびえるけど遊ぼう!
 - 一時帰宅したしもう外出たくない
 - 満足して一時帰宅してしまった
 - 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
数年ぶりに気持ち悪い思いました
 - 気持ち悪い客しか来ない
 - 何の話かわからんけど恥ずかしい
昨日は本気で疲れすぎてぶっ倒れてた
とりあえず、まあこれからよろしくお願いします
また出会った時は協力してね
 - 駅で先生に出会って気まずい

は若者言葉を言い換えた標準語表現部分を示す。

日露翻訳では 18 文のうち 11 文が正解だった (図 3)。Excite 翻訳と Yandex 翻訳は翻訳成功になった文はそれぞれ 3 文と 2 文しかなく、精度は低い。これに対して、Google 翻訳は全 18 文成功した中に正解 11 文も含まれており、最適翻訳文はほぼ Google 翻訳の成果である。

日中翻訳の場合は 18 文のうち 16 文が正解、残り 2 文が意図推定可能となった (図 3)。Google 翻訳の精度は日中翻訳でも一番高く、翻訳成功 13 文のうち 10 文が正解だった。百度翻訳の精度は Google 翻訳よりやや低く、翻訳成功 11 文のうち 8 文が正解であった。Excite 翻訳は正解文 3 文、意図推定可能 7 文となったことから、Excite 翻訳の

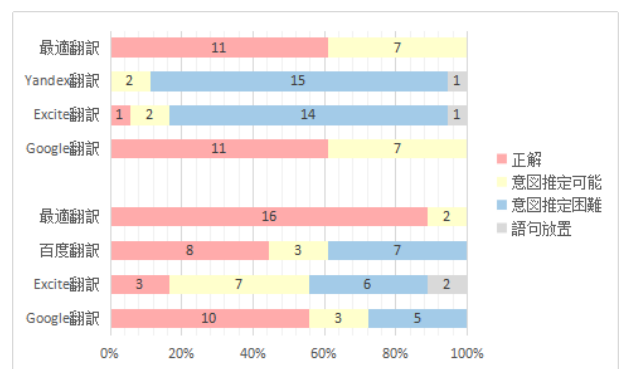


図 3 翻訳成功文の翻訳結果 (上:日露翻訳, 下:日中翻訳)

日中翻訳での精度は日露翻訳より高いと考えられる。

表4の18文中、日露翻訳と日中翻訳の双方で正解となった文が9文あった(表4○印)。この9文の各翻訳器での日露翻訳の結果(図4)では、Google翻訳のみ全9文正解と

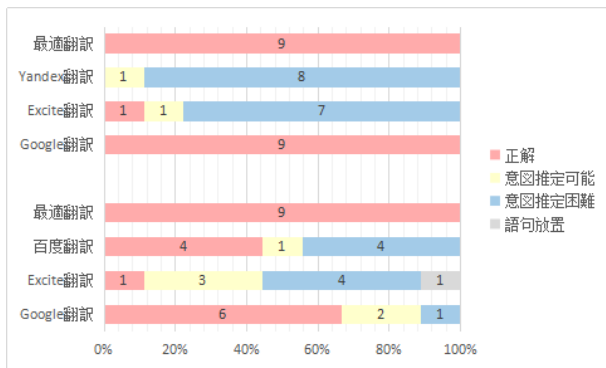


図4 双方で正解した9文の各翻訳器の出力(上:日露, 下:日中)

なった。Excite翻訳で正解となった文が1文、Yandex翻訳は正解文を出すことができなかったが1文意図推定可能の文を出した。日中翻訳の場合、Google翻訳の精度が一番高く、9文のうち正解文が6文あり、1文のみ失敗した。百度翻訳の精度はGoogle翻訳よりやや低く、9文のうち4文が正解になって、4文が翻訳失敗した。Excite翻訳は日中でも正解文1文を出すことができ、日露翻訳より2文多く意図推定可能の文を出した。そして、図1、図2、図3と図4から、Google翻訳は2言語への翻訳でほぼ一致した精度を挙げることができるだけでなく、1つの例文に対し2言語への翻訳でも翻訳成功を出せる可能性は高いとも考えられる。

図5に、若者言葉を標準語表現へ言い換えた結果翻訳に成功した例を挙げる。図5(a)では、「リア充」が「リア(后

若者言葉での翻訳文	言い換えでの翻訳文
(a) リア充になる計画失敗だ 是成为后面充的计划失败	現実の生活が充実している人になる計画失敗だ 是成为现实生活充实的人的计划失败
(b) これ年末ガチでやばい 这个年末gachi危险	これ年末本気でやばい 这个年末当真危险

図5 言い換えによって翻訳に成功した例

面: 後ろ)」と「充」に分けて翻訳されたため意味推定困難な翻訳文が出力されていたのに対して、これを言い換えることで「现实生活充实的人(现实生活は充実している人)」とそのまま訳され、意味のわかる文となった。図5(a)では中国語文の変化は「リア充」部分のみで、その前後は同一である。図5(b)は「ガチ」が翻訳できず「gachi」とローマ字で出力された部分が、「当真(本当に)」に置き換わっている。

3.3 話し言葉文の翻訳の問題点

中国語の場合言い換えにより正解率は100%に向上した(図1)が、ロシア語の場合は60%の翻訳成功に留まった(図2)。そこで、若者言葉の標準語表現への言い換えだけでは翻訳精度を向上させることができない原因を検討する。

日中翻訳では言い換えにより翻訳に成功した(図5)のに対して日露翻訳では同じ言い換えでも成功しなかった例を図6に挙げる。図6(a)では、「リア充」が「リア(後ろ)」と

若者言葉での翻訳文	言い換えでの翻訳文
(a) リア充になる計画失敗だ Это план не смог заполнить тыл (この計画は後部を埋めることができませんでした)	現実の生活が充実している人になる計画失敗だ Планирование не может стать человеком, который полон реальной жизни (計画は、現実の生活に満ちている人にはできません)
(b) これ年末ガチでやばい В этом году учитель и горячие (今年は先生と熱い)	これ年末本気でやばい Этот конец опасным (この終わりは危険です)

図6 中国語とロシア語で結果の異なった例

「充」に分けて翻訳されたために意味推定困難な翻訳文が出力されたこと、言い換えによって「現実の生活が満ちている人」と翻訳が変化したことは日中翻訳と同様だが、日露翻訳では文の構成要素の把握に誤りがあり、結果的に意味の異なる文となっている。図6(b)は「ガチ」の部分が言い換え前だけでなく言い換え後でも翻訳結果から抜け落ちてしまっており、「本気で」を翻訳することができていない。

日露翻訳(図2)では、標準語表現への言い換えを行った後の文でも最適翻訳文が翻訳失敗に留まる文が30文中12文ある。この12文の各翻訳器の翻訳結果(図7)を見てみると、日露翻訳では3種類の翻訳器がすべて翻訳に失敗し

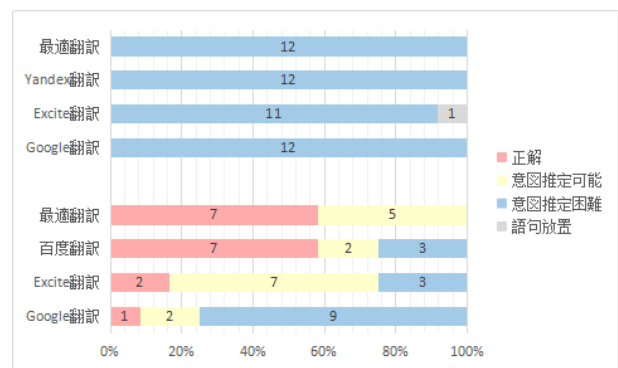


図7 日露翻訳失敗12文の各翻訳器の出力(上:日露, 下:日中)

たが、同じ文について日中翻訳では12文すべてについて最適翻訳文は翻訳成功となり、うち7文が正解となった。この12文についてはGoogle翻訳は日露翻訳でも日中翻訳でも精度は低く、成功は日中翻訳での3文のみだった。日中翻訳の場合、Excite翻訳と百度翻訳は12文のうち9文が成功したが、百度翻訳はExcite翻訳より5文多く正解文を出すことができている。日本語から対応言語へ翻訳する際、百度翻訳の日本語処理アルゴリズムの方が精度が高い

と考えられる。

図7に挙げた翻訳失敗は、若者言葉を標準語表現へ言い換えた日本語文に対する翻訳の結果である。若者言葉を含む文の翻訳に失敗する理由としては、若者言葉が辞書に記載されておらず対訳語が得られない可能性が挙げられる。この場合、Google翻訳のように文全体をニューラル翻訳するような手法でも、若者言葉をほかの語句と混同するなどして翻訳に失敗する可能性が高い。

そこで本稿では、ロシア語で最適翻訳を選択しても失敗になった翻訳文12文(図7)について失敗の原因を調べ、表5の5種類に分類した。本稿では表5を話し言葉失敗の

1 語句選択誤り	3 構文要素誤り	4 慣用表現欠如
2 語句削除	(1) 語句補充誤り	5 語句放置
	(2) 語順誤り	

原因の分類と捉え、ロシア語だけでなく中国語の例も挙げて説明する。ただし、5分類のうち慣用表現欠如と語句放置については、この12文に対する中国語翻訳の出力に例がないため、ロシア語のみ例を挙げる。

3.3.1 語句選択誤り

格変換の間違いや単語を文意と異なる意味の語句に翻訳するなど語句選択に誤りがある場合、これを語句選択誤りと考え。例えば図8(c)では「いて出会わない」が「会が出ることがある」と訳されており、文意が変化している。図8(d)では原文の「恥ずかしい」は「怪しい」という意味の語に翻訳されている。

(c) 同じ建物にいて出会わないって逆に凄い 在同一栋建筑上会有会出, 反而很厉害 (同じ建物で会が出ることがあります、かえってすごいです)
(d) 違っていたら恥ずかしいけど Странно, если все по-другому (それが異なるのであるならば、それは怪しい)

図8 語句選択誤りの例

3.3.2 語句削除

翻訳されるべきの原文の一部が翻訳されていない場合を語句削除とする。例えば、図9(e)では、原文の原因を指す「から」が翻訳文では抜け落ちている。図9(c)では原文の「凄い」は翻訳されていない。

(e) とりあえず、まあ人多すぎるから会場から出る 无论如何, 有太多的人从场地出来 (いずれにしても、あまりにも多くの人が場所から出てくる)
(c) 同じ建物にいて出会わないって逆に凄い Том же здании, что собрание не имеет обратного, так что я (会議には逆のことがないのと同じ建物なので、私は)

図9 語句削除誤りの例

3.3.3 構文要素誤り

構文要素誤りはさらに語句補充誤りと語順誤りの2種類に分ける。

3.3.3(1) 語句補充誤り

原文で主語、助詞などの構文要素が省略され動作主が不明の場合に翻訳器が文意に沿わない構文要素を補充し、翻訳に失敗する事例を指す。例えば図10(f)の例では、原文で主語(第二人称)が抜けているため翻訳に際してこれを自動的に第一人称で補った誤りである。図10(g)の原文は「他人に見られることを恥ずかしい」との意味と捉えるのが妥当だが、「他人に」に当たる部分が抜けており、これを第一人称で補った結果文意が変わっている。

(f) とりあえず、まあこれ守ってれば痩せれるから 无论如何, 如果你保持这个好, 我会减肥 (とにかく、これを保つなら、私は体重を減らすでしょう)
(g) 自分のパソコンのデータ見られるのとか超恥ずかしいぜ Смущает видеть данные моего персонального компьютера (私のパーソナルコンピュータのデータを見るのは恥ずかしいです)

図10 語句補充誤りの例

3.3.3(2) 語順誤り

形態素解析または係り受け解析で構文要素の認識に失敗することで語順が入れ替わり意味が通らなくなることを語順誤りとする。例えば図11(a)の例では、日中翻訳でも日露翻訳でも語順が乱れていて、適切に翻訳できていない。図11(a)は、元の若者言葉(「リア充」)が名詞1個だった

(a) 現実の生活が充実している人になる計画失敗だ 计划成为一个真正的人是失败的 (本当の人になるのを計画するのは失敗的である) Планирование не может стать человеком, который полон реальной жизни. (計画は、現実の生活に満ちている人にはできません)

図11 語順誤りの例

のに対して言い換えに用いた標準語表現(「現実生活が充実している人」)が文を含有するような長い名詞句であり、このような複雑な表現への言い換えが語順誤りを誘発した可能性がある。

3.3.4 慣用表現欠如

慣用表現欠如は、翻訳先言語に対応する表現がない日本語文の翻訳に失敗することを指す。図12(h)の例では「嬉しくなる」といった意味の「気持ち上がる」という日本語の言い回しが、直訳されてもおらず、意味推定ができなくなっている。逆に、「気持ち上がる」の部分を直訳した

(h) 最近お部屋めっちゃキレイだから気持ち上がる В последнее время номера так красиво, потому что они очень красивые (最近、部屋は非常に美しいのでとてもきれいです)
--

図12 慣用表現欠如の例

翻訳器は、出力結果が意味推定可能となった。こういった慣用表現をさらに平易な標準語表現に言い換えることができれば、翻訳精度の向上が期待できる。

例えば、「これからもよろしくお願ひします」という日本語表現にはロシア語ではこれに対応できる文が無い。こう

いった慣用表現は意味合いが合うように翻訳することは難しく直訳されることが多いため、標準語表現への言い換えにあたっては、慣用表現を避けることも考慮すべきである。

3.3.5 語句放置

語句放置とは、語句レベルで対応する翻訳語句を決められずローマ字表記などを含んで出力する失敗を指す。例えば図13(f)の例では、「瘦せれる」の部分で認識できずローマ字表記に置き換えられており、翻訳文として成り立っていない。図13(f)の例の失敗原因はら抜き言葉で、「瘦せれ

(f) とりあえず、まあこれ守ってれば瘦せれるから
От Семьянкегеги в настоящее время хорошо сохранился
только
(Semerankereruから現在はよく保存されている)

図13 翻訳失敗の例

る」の部分で「瘦せられる」に修正すると語句放置は解消する。ただし、この例ではら抜き言葉を修正しても同じ翻訳器での翻訳結果は「しばらくの間、とにかく薄くなる場合それ」という意図推定困難なもので、翻訳失敗である。

3.4 日中翻訳と日露翻訳の相違点

図2で翻訳に失敗している12文について、主な原因を検討するため複数の翻訳器それぞれの誤り原因の数を調べた結果を図14に示す。図14では、ある1文が複数の誤り

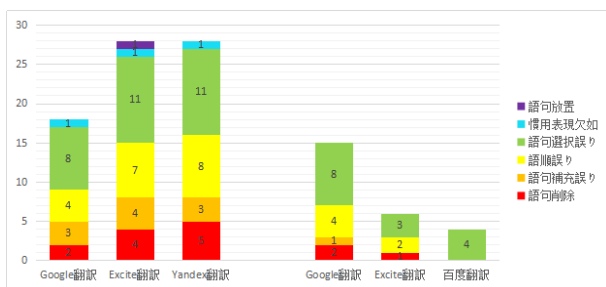


図14 日露翻訳(左)と日中翻訳(右)の誤り原因

原因を含んでいる場合、それらは独立に数えた(例えば、ある1文が3種類の誤りを含んでいる場合、それぞれの誤りについて1文と数えた)。そのため誤り文の合計数は12文を超える場合がある。

図14を見ると、5つの誤り原因のうち、日中翻訳、日露翻訳の双方で語句選択誤りが特に多く、二番目は構文要素誤り(語順誤り)である。図14の結果から、中国語にとってもロシア語にとっても、構文要素は2つの言語へ翻訳するのに重要だということがわかる。これは用いた翻訳器のアルゴリズムの良し悪しと大きな関係があると考えられる。

日露翻訳で用いた3つの翻訳器ではGoogle翻訳の精度が一番よく、1つの訳文に対し複数の誤りがある例は他の2つの翻訳器よりは比較的少ない。Google翻訳は中国語の場合でもほぼ一致した精度を得た。このことから、Google

翻訳は多言語間の翻訳を行う際、精度を保てることがわかる。中国語への翻訳で用いた3つの翻訳器のうち精度が一番良いのは百度翻訳である。Excite翻訳器の精度は日露翻訳よりも日中翻訳の方が高い。

4. まとめ

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に比べて、若者言葉を標準語表現へ言い換えてから翻訳した場合は日中翻訳でも日露翻訳でも明らかに精度が向上しており、若者言葉を言い換えることで翻訳が容易になること、正解率の向上に寄与することを確認した。

本稿では、若者言葉を含む話し言葉文の翻訳精度の向上のため、日中翻訳と日露翻訳とを対象に複数の翻訳器について問題点を検討した。その上で、話し言葉の翻訳で問題となる点は5種類に分類でき、特に構文要素の同定が重要と考えられることを示した。

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] 何 婉瑩, 藤田 和成, 浅沼 爽汰, 王 美キ, 郭 秀景, 中島 浩平, 町田翔, 延澤 志保, “日中翻訳のための若者言葉の対訳分析,” 情報処理学会第80回全国大会, vol.2, pp.323-324, 2018.
- [3] 藤田 篤, 乾 健太郎, “語訳文を利用した普通名詞の同概念語への言い換え,” 言語処理学会第7回年次大会発表論文集, pp.331-334, 2001.
- [4] 松本 和幸, 土屋 誠司, 芋野 美紗子, 吉田 稔, 北 研二, “感性を考慮した日本語俗語の標準語変換,” 人工知能学会論文誌, vol.32, no.1, pp.1-12, 2017.
- [5] 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.
- [6] 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>.
- [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://bosesound.blog133.fc2.com/>.
- [9] Twitter: <https://twitter.com/>.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 日本政府観光局 2017 年度の国籍/月別訪日外国人統計データ: <https://www.jnto.go.jp/jpn/>.