

翻訳困難例文の機械翻訳と評価

横山 晶一（山形大）、徳永 信治（シャープ）、亀田 雅之（リコー）、
内尾 淑美（富士通）、芦崎 達雄（J I C S T）、野田 庸男（C S K）、
相澤 弘（十印）、中 英康（N E C）、田中 穂積（東工大）

アジア太平洋機械翻訳協会 (Asian-Pacific Association for Machine Translation, 略称AAMT) 例文評価研究会は、機械翻訳システムにおける種々の問題点を、自然言語処理の観点から見直し、研究していくという目的で活動を行っている。1992年から1993年にかけて、日本語の翻訳困難例文を収集、分析した（別稿）。本稿では、翻訳困難例文と人手によって作ったその言い換え例を機械翻訳システムにかけた結果と、その分析について述べる。いずれの例文も、前処理や未定義語の登録を行わずに機械翻訳している。また、翻訳困難例文そのものについては、人間による翻訳も参考とした。また、これらの例文について、辞書や構文レベルなどで、ユーザに対して自動的に警告が出せるかどうかについても検討した。

Machine Translation and Evaluation of Japanese Sentences Difficult to Translate

Shoichi YOKOYAMA (Yamagata Univ.), Shinji TOKUNAGA (Sharp Co.),
Masayuki KAMEDA (Ricoh), Yoshimi UCHIO (Fujitsu), Tatsuo ASHIZAKI
(JICST), Tsuneo NODA (CSK), Hiroshi AIZAWA (Toin), Hideyasu NAKA
(NEC), and Hozumi TANAKA (Tokyo Inst. Techno.)

Sentence Evaluation Committee in Asian-Pacific Association for Machine Translation (AAMT) was established for studying various issues in machine translation from the viewpoint of natural language processing. In 1992 and 1993, we collected sentences difficult to machine translate (written in another paper of this SIG). This paper describes the result and analysis of machine translation of these sentences and their paraphrasing by human. Every sentence is machine translated without pre-editing nor registration of unknown words. Sentences difficult to machine translate are referred to their translation by human. We also discussed whether automatic warning of such sentences in lexical or syntactic level to users is possible or not.

1. はじめに

アジア太平洋機械翻訳協会（会長：長尾真 京都大学教授）(Asian-Pacific Association for Machine Translation, 略称AAMT) 例文評価研究会（委員長：田中穂積 東京工業大学教授）は、機械翻訳システムにおける種々の問題点を、自然言語処理の観点から見直し、研究していくという目的で活動を行っている。

本稿では、この委員会が 1992 ~ 1993年にかけて行った活動の一端を報告する。この期間には、機械翻訳システムの評価と改善につながるものとして、まず、翻訳が困難な例文を選び出して分類し、その性質を検討した。また、例文の言い換えを人手で行った。これらの例文と言い換え例を、前処理なしで機械翻訳システムにかけ、その結果について検討を行った。本稿では、機械翻訳の結果とその評価について述べる。例文の選定とその分類、言い換え例の作成については別稿 [1] にゆずる。

2. 例文、言い換え文の機械翻訳

選定された翻訳困難例文（日本語）と、それに対する言い換え文を、前処理や辞書登録なしで日英機械翻訳システムにかけた。また、例文については、人間による翻訳を行い、評価、検討の際に参考にすることにした。人間による翻訳は、意訳されていることが多いので、これを基準にすることはせず、あくまでも参考にとどめた。

図 1 に、「自分でラジオが直せますか」という例文に対する機械翻訳例と、その言い換え文 ((1A) から (1H) という文番号で示す 8 種類の文) に対する機械翻訳の例を示す¹⁾。(1) で # で示した "Can you repair a

radio by yourself? " という文が、人間による翻訳である。また、その他の翻訳文は、市販または研究開発されている日英機械翻訳システムによる訳文である。

異なるシステムでも出力文が同じ場合には、一つにまとめてある。したがって出力文の数は必ずしも使用したシステムの数とは一致しない。この例にはないが、辞書が未登録のために日本語がそのまま出力された場合にはそのままの形で残してある。解析や出力が不完全な場合でも、文の形をなしている限りはできるだけ出力として採用するようにした。また、目的語や主語を補つて訳す場合には、システム固有の記号を用いず、SUBJ, OBJ などで示すようにした。

図 1 では、構文的に類似のものをなるべく近くにまとめるようにしてある。(1) の例では、助動詞として "can" を用いるかどうか、"radio" か "a radio" か "the radio" か、動詞では "mend", "correct", "repair" のどれを用いているか、再帰的な部分を "myself", "oneself" のどちらを用いているかなどの違いがなるべく把握できるような配列にしているつもりである。

言い換え例は、機械翻訳が容易になるようになるべく元の文の意味を変化させないようにして、人手で作成した例文である。図の例で見ると、たとえば英語の直訳調である (1B) 「あなたはあなた自身でラジオが直せますか」では、(1) の人間訳に比較的近い翻訳文が得られているが、すべての言い換え文が必ずしも機械翻訳を容易にする方向に向かっている訳ではなく、たとえば (1D) 「あなたはあなた自身の手でラジオが直せますか」といった文では、「手で」という語句が入った分、処理が複雑になり、訳のバリエーションが広がっている傾向が見られる。

¹⁾ 全例文、言い換え文に対する訳例については [2] 参照。

(1)自分でラジオが直せますか。

- Can a radio be mended for oneself?
- Can radio be corrected personally?
- Can the radio be mended for myself?
- Can the radio be for myself mended?
- Can radio repair OBJ with self ?
- Is it possible to repair a radio by it?
- Does a radio correct it by a self?
- Does the radio set convert it by the oneself?
- # Can you repair a radio by yourself?

(1A)あなたは自分でラジオが直せますか。

- Can you mend the radio yourself?
- Can you mend a radio for oneself?
- Can you for myself mend the radio?
- Can you repair radio with self ?
- Can you repair a radio from oneself?
- Can you repair a radio by you?
- Can you repair a radio in me?
- Does the radio set convert you by the oneself?
- Does a radio cure you by yourself?
- Can you correct radio personally?

(1B)あなたはあなた自身でラジオが直せますか。

- Can you mend the radio by you?
- Can you mend a radio in yourself?
- Can as for you you repair a radio from oneself?
- Can you repair radio with self ?
- Can you repair a radio by yourself?
- Does the radio set convert you by the you oneself?
- Does a radio cure you by you oneself?

(1C)あなたは自分の手でラジオが直せますか。

- Can you mend the radio by your hand?
- Can you mend a radio in the hand of oneself?
- Can you repair a radio with my hand?
- Can you repair a radio with your hand?
- Can you repair radio we are manually self ?
- Do you correct it about radio at the means of the oneself?
- Does a radio cure you by your hand?
- Can you correct radio by your hand?
- Can you repair a radio from (your) help?

(1D)あなたはあなた自身の手でラジオが直せますか。

- Can you mend the radio by your own hand?
- Can you mend a radio in your own hand?
- Can you repair radio self manually ?
- Can you repair a radio from (your) own help of you?
- Can you repair a radio with your own hand?
- Do you correct it about radio at the means of the you oneself?
- Does a radio cure you by your oneself hand?

(1E)あなたはラジオを直せますか。

- Can you mend the radio?
- Can you mend a radio?
- Can you repair radio ?
- Can you repair a radio?
- Do you repair radio?
- Do you repair a radio?
- Can you correct radio?

(1F)あなたはラジオが直せますか。

- Can you mend the radio?
- Can you repair radio ?
- Can you repair a radio?
- Does the radio set convert you?
- Does a radio cure you?
- Can you correct radio?

(1G)あなたはラジオを修理できますか。

- Can you repair the radio?
- Can you repair the radio set?
- Can you repair radio ?
- Can you repair a radio?
- Can you fix radio?

(1H)自分でラジオを直すことができますか。

- Can the radio be mended by me?
- Can a radio be mended for oneself?
- Can you repair radio with self ?
- Can (you) repair a radio from oneself?
- Can radio be repaired by the oneself?
- Can a radio be repaired by a self?
- Can a radio be repaired in me?
- Can radio be corrected personally?
- Is it possible to repair a radio by it?

図1 「自分でラジオが直せますか」と言い換え文の機械翻訳例

3. 翻訳結果の評価と分析

例文、言い換え文の翻訳結果について、日本語文と英語文とを比較する形で検討、分析を行った。分析の際の分類項目は、例文の分類項目とは異なっている。この分類にしたがって分析を行い、言い換えのための対策を検討した（以下の分析結果の中で「対策」として示す）。対策にしたがって試みた言い換え例も示した。さらに、翻訳困難文へ警告が出せるとすると、現在の研究レベルでどの程度可能かということについても検討した。この可能性とは次の通りである。

翻訳困難文への警告の可能性

レベル1 (*) : 表層で警告が出せる。
レベル2 (**) : 品詞列レベルで警告が出せる。
レベル3 (***) : 警告を出すのは難しい。

なお、レベル3については、「表層あるいは品詞レベルでは警告を出すのは難しい」という意味と「現在の自然言語処理、機械翻訳のレベルからいって警告不能」という意味とが混在している。このレベルは、コメント欄で示し、コメントを付け加えた。以下に分析の一部を示す。項目はすべてをかかげるが、紙数の関係で、例文は一部を、また、説明、言い換え例、コメントについては少數例を示すことにとどめる。文番号は、最初の英大文字と数字が例文の分類項目、ハイフンの後の数字、英小文字がその項目での例文番号、最後の英大文字が言い換え文を示す。

- (あ) 日本語解析の問題点
 - (あ1) 構文に起因するもの
 - (1) 「…は～が」文

(a) 「AはBが」 → 「AのBが」または「AのBは」となるもの。

(B5-1) 妻は顔がきれいだ。

[対策] 所有関係、または部分全体関係に適合するものの言い換え。

[言い換え例] 妻の顔はきれいだ。

[コメント] * * : 「…は～が」パターンに加えて、述語部分が形容詞、形容動詞に限って検出するには、品詞情報が必要。

(b) 「AはBが」が上記のようにならないもの。

(B5-2) この事故は君の不注意が原因で起こった。

[対策] 使役文に書き換える。

[言い換え例] この事故は君の不注意が起こした。

(B5-5B) この報告は、彼が訂正する。

[説明] 形は「…は～が」文であるが、曖昧性なく翻訳が成功する例。

[コメント] 述語の格情報で対処可能。

[全体に関するコメント] * : (a) タイプか(b)タイプかの判断は難しい。「AはBがC」でAとBの意味関係だけでは判断できない。Cの属性、BとCとの関係にもよる。

(2) 「…は～は」文

(A14-4) オペランドは必ずしも指定する必要はありません。

(3) 「…は～を」文

(B5-3) セミナーは日程を一日延長する。

(4) 連体修飾・埋め込み文における係り受けの曖昧性

(A13-3a) 放送局へ行く道を教えてください。

(A13-3b) 放送局へ行くバスを教えてください。

(5) 連用修飾における係り受けの曖昧性

(A14-2) アメリカの希望によって、
アメリカ国内で生産し、販売す
るという動き。

(6) 並列構造（長文の項も参照）

(A2-1c) 日本の動物と植物の歴史。
[対策] 括弧を使用して係り受
け関係を明確にする。

(7) 平叙文の形をした命令文・禁止文

(A14-4) オペランドは必ずしも指定
する必要はありません。

[言い換え例] オペランドを必ず
しも指定する必要はありません。

(8) 引用・述語のスコープ

(A14-2) タバコは人体に有害だと言
われている。

[対策] 引用の範囲を括弧で明示。

[言い換え例] [タバコは人体に
有害だ] と言われている。

[コメント] * * :

(9) 長文（並列の項も参照）

(A9-1) 組織における実際の作業は、
複数人による共同作業という形
態をとることが多いにもかかわ
らず、そのような共同作業を支
援する環境はあまりなく、個人
の生産性向上に重点が置かれて
いるのが現状である。

[対策] 読点を手掛けりにして
文を分離することを考える。文
脈理解の項も参照。上記の例で
は、読点の前の「にもかかわら
ず」を手掛けりにして、次の一
うな言い換えをする。一般的に
は、読点の前の接続表現、連用
中止法などをキーにする。

(10) 埋め込み句の「が」と主語の「が」

(B11-3cA) 彼女が先に着た服はいや
だ。

(あ2) 意味に起因するもの

(1) 多義性

(B11-7) 車が家になる。

[対策] 意味関係を明示する。

[言い換え例] 車は (or を) 家と
して使うことができる。

車に住んでいる。

(B2-2) データを書き出しファイル
に保存する。

(2) 数量表現

(B2-1a) 彼は一部修正した。

[対策] なるべく一義に決まるよ
うに言い換える。

(D2-1) 牛乳を一日に三回飲む。

(あ3) 品詞に起因するもの

(1) 和語動詞

(A11-2) 危険な液体には色がつけて
ある。

[対策] 漢語動詞「着色されてい
る」などに言い換え。

慣用表現に対する辞書の充実。

[言い換え例] 危険な液体は着色
されている。

危険な液体は色付けしてある。

(2) サ变动詞

(B5-9D) 私は週末は決まってテニス
をする。

[説明] 「行為名詞+する」と
「行為名詞+をする」の違い。

[対策] 目標言語との変換辞書の
充実。

(3) 助詞「で」の曖昧性

(B11-6d) 満員のバスで会社に行く。

[説明] 手段を表す「で」は、し
ばしば接続助詞（学校文法では

助動詞「だ」の運用形)と間違われる。

[対策] 「で」を明示的な言葉で置き換える。

[言い換え例] 満員のバスに乗つて会社に行く。

(4) 助詞「て」の曖昧性

(B11-5C) 彼は株で儲けて家を建てた。

(B11-5D) 彼は株でお金を儲けて家を建てた。

[説明] 結果を表す「て」であるが、並列と誤解釈される。

[対策] 「儲けた結果」、「儲けたので」と言い換える。

(5) 助詞「と」の曖昧性

(B11-4G) 山本と呼ばれる人がやつて來た。

[説明] 「と」を並列の助詞と誤解釈することがある。

[対策] 「呼ぶ」、「呼ばれる」の辞書に共起格として「と」を記載。

(6) 助詞「の」の曖昧性

(B6-1) この種の推論プロセスの具体的なメカニズムは現時点では明らかにされていないが、この推論機構の解明が、自然言語の柔軟な意味理解のモデル化の中心的な問題となる。

[対策] だらだらと続く文を書かない。反意接続でない「が」の項も参照。

(7) 助詞「によって」「のために」の曖昧性

(B11-1a) 勉強のために、目が悪くなつた。

(B11-1aA) 勉強によって、視力が低下した。

[説明] 原因を表す「によって」、「のために」は曖昧な場合がある。2(2), 2(6)も参照。

[対策] 原因は「によって」の方がよい。

(8) 可能助動詞「れる」「られる」

(B2-3a) まだ着られた服を捨てた。

(9) 反意接続でない「が」

(B6-1) この種の推論プロセスの具体的なメカニズムは現時点では明らかにされていないが、この推論機構の解明が、自然言語の柔軟な意味理解のモデル化の中心的な問題となる。

(10) 補助動詞、アスペクト

(C9-1) 水を汲みに行って来てください。

(あ4) 表現・文脈理解に関するもの

(1) 口語的表現

(B11-2) 君があの絵が好きだなんて驚きだ。

(2) 指示代名詞による照応

(B11-4) 今夜は私の番なので、それが悲しくて泣いているのです。

(3) 敬語表現

(B11-3b) 天皇陛下が着られた服を拝見した。

(4) 慣用表現

(B5-13) 佐藤は佐藤なりの仕事しかできない。

(5) 二重否定

(B5-12aD) この病気は注射しなければ直らない。

(6) 文脈表現

(C42-2b) 息子には、現金を考えている。

[説明] 前の文「妻には、遺産として自宅と土地を残す」を受け

ている。

[対策] 複数の文を扱う処理が必要。「現金を相続させること」を補うことも考える。

(7) 特殊な表現

(A9-4) 回路が短絡していると、ヒューズが溶断することはもちろん、場合によっては機器を損傷する恐れもあるので、電源を投入する前に回路抵抗を測定することを怠ってはいけません。

[説明] 微妙なニュアンスが入ったもの。

[対策] 簡潔な表現に言い換える。「溶断することがある。さらに、…もある」

(い) 目標言語（ここでは英語）との対比における問題点

(1) 日本語における主語の省略・欠如・転換

(B11-6b) 満員のバスに乗って遅刻した。

(B11-6c) 満員のバスに乗って恥をかいた。

[説明] 主語は「私」などが考えられる。

[対策] 主語として「私」、「我々」、「彼」などを補う。どれを補うかを自動的に判断するのは困難。なお助詞「て」の曖昧性の項も参照。

(2) 英語における無生物主語

(B11-1aA) 勉強によって、視力が低下した。

(B11-1aA) 勉強によって、視力が低下した。

[説明] 原因を表す「によって」、「のために」は英語で無生物主

語となることがある。

[コメント] 変換辞書に明記する。

(3) 発想の違いによる特別な表現。

(C1-3) 彼女には子供がいる。

(4) 目的語の欠如

(A13-3a) 放送局へ行く道を教えてください。

[説明] 英語では、教えられる人(me, us)の明確化が必要。

[対策] 目標言語の辞書にある必須格が原言語に欠けている場合のdefault語の必要性。

(5) 再帰代名詞

(A11-1) 自分でラジオが直せますか。

(6) 所有代名詞の欠如

(B11-1a) 勉強のために、目が悪くなつた。

(7) 原言語と目標言語との対応が1対多になる場合

(A13-3a) 放送局へ行く道を教えてください。

(C1-5) 道が汚れた。

[説明] 上では、道→way（方法、手段）、下では、道→road（物理的な道路）。

[対策] 共起する動詞との対応関係の記述が必要。

(8) 比較表現

(A14-5) ユーザは置換モードよりも挿入モードの方を高く評価する。

[説明] 比較表現の訳がうまく出ない。

[コメント] 「方を」「より」といった比較表現を捉える辞書を作る。

* : 同一文内の
「より」、「方」の共起を検出。

(9) 被害の受け身

(A34-1) 彼女は鳥に逃げられた。

[対策] 多少ニュアンスは異なるが、自動詞構文に言い換える。

[言い換え例] 彼女の鳥が逃げた。

(10) 共起関係の異なり

(B11-3D) 私がいない間、誰かが私を訪ねてきたらしい。

(B11-3E) 私がいなかつた時、誰かが私を訪ねてきたようです。

[説明] 「いない」は、英語では場所を表す格と必然的に共起する。

[対策] 「その場所に」などを補うか、英語生成過程で場所格を付ける。

(11) 日本語定型文

(C1-1a) 拝啓、陽春の候、貴社ますますご盛業の趣、大慶に存じ上げます。

[コメント] 無理に翻訳せず、対応する英語定型文に置き換える。

(う) 目標言語生成に関する問題点

(1) 冠詞の問題

[対策] ネイティブによる後処理。

(2) 前置詞

(B5-9) 私は週末はテニスに決めている。

[説明] 副詞表現に対する英語の前置詞。「…は～は」文の項も参照。

[対策] 英語生成辞書をしっかりと作る。

(3) 時制の一貫

(B2-3a) まだ着られた服を捨てた。

[説明] 「着られる」と「着られた」で時制が異なることがある。可能助動詞の項、主語の欠如の項も参照。

[対策] 可能を明示した表現に改

める。

4. おわりに

翻訳困難例文をやや恣意的な観点から選んだために、主として科学技術論文を翻訳対象としている多くの機械翻訳システムにはふさわしくない作例がいくつか指摘された。この反省に立って、現在、実際の科学技術論文の文例からいくつかを選択し、分析、機械翻訳を行っている。また、人間による翻訳がどちらかといえば意訳で、機械翻訳の分析に必ずしも参考にならないことから、現在は、機械翻訳出力文を人間によつて後処理をした結果も比較対照に加えることを考えている。

また、例文の分類項目、翻訳文と比較対照したときの分析項目についてもさらに検討を進めていく予定である。

謝辞

本委員会を支えて下さるAAMT運営委員会ならびに事務局（事務局長 星野 穎男氏）の方々に感謝します。また、分析の途中まで委員、オブザーバーとして本委員会の活動に参加された方々に感謝します。さらに、この分析結果をお送りし、ご多忙のところ有益なコメントを数多くいただいた（残念ながら本稿にはほとんど収録していない）、石綿 敏雄 日本女子大学教授、成田 一 大阪大学助教授、水谷 静夫 東京女子大学名誉教授（計量計画研究所理事）の3氏に感謝の意を表する。

参考文献

[1] 横山、熊野、望月、植村、川越、時岡、安藤、田中：翻訳困難例文の収集と分類、本研究会資料101-5 (1994)

[2] 研究成果報告書（93年度版）、アジア太平洋機械翻訳協会(1993)