AI翻訳にも限界がある

AI Translation has its limitations

大阪大学　名誉教授
成田　一

英日対照構造論・機械翻訳・言語教育 / 外国語習得論専攻。大阪大学功績賞受賞。著書『パソコン翻訳の世界』（講談社）、『日本人に相応しい英語教育』（松柏社）ほか、編著『こうすれば使える機械翻訳』（バベルプレス）、『英語リフレッシュ講座』（大阪大学出版会）ほか、共著『英語教育徹底リフレッシュ』（開拓社）ほか。『翻訳技術の言語的な基盤』（51 ヶ月連載：バベルプレス）、『総合的な翻訳による英語教育』（50 ヶ月連載：バベルプレス）、『Translators』（日本翻訳協会）、『Cat』（アルク）の連載、『SPA』（扶桑社）、『英語教育』（大修館書店）、『新英語教育』（三友社出版）などの雑誌記事、英文テキスト、新聞コラム、論文など多数。英語教育総合学会会長。

本稿では、かつてのルールベースの機械翻訳に求められ、現在でもある翻訳企業において実施されている「前処理」を取り上げ、その多くが近年の AI 翻訳には必要ないことを明確にする。2016 年以降、『Google 翻訳』、『DeepL 翻訳』、『VoiceTra 翻訳』など AI による深層学習（Deep Learning）により極めて高度な翻訳が可能なニューラル翻訳プログラムが開発、公開されたが、これはあくまで言語レベルの処理だ。人間のように意味を理解した翻訳ではない。このため、修飾関係の曖昧性を解消したり、発話の意図を汲んだ意訳はできない。そうした高度な意味処理は人間が行うにしても、AI 翻訳に委ねて良い文書は膨大だ。

1　前編集は必要か？

日本で大手電気通信系メーカ各社において市販の機械翻訳が開発されたのは 1980 年代後半だが、この頃は翻訳能力が低く実用段階に達していない初期の機械翻訳が苦手とする文は「機械翻訳が訳せる文に予め編集して機械翻訳に入力」していた。その後、機械翻訳の翻訳能力向上に伴い、前編集を行う企業は減ってきたが、自社の翻訳ソフトを活用するにあたって前編集を今も行っている企業がある。多言語翻訳サービスや翻訳ソフト『PC-Transer』の開発を行ってきた「クロスランゲージ社の前編集の説明」を（表記の一部を太字化、下線を引くなど若干編集して」）見てみよう。

2　「クロスランゲージ社の前編集の説明」

ここでは、日→英翻訳に欠かせない「前編集」のルールについて説明します。

コンピュータで翻訳 (日本語→英語) するときは、「前編集」の作業が必須です。では、なぜ前編集が必要なのでしょうか。それは、「コンピュータは機械的に翻訳する」だけだからです。原文の日本語以上の英語を生成できないのです。つまり、日本語がわかりにくかったら、わかりにくい英語として翻訳されてしまうのです。

コンピュータは人間と違って、行間を読んだり、文章の前後関係を考慮しません。翻訳対象となる原文だけを見て解析します。したがって、翻訳結果の善し悪しは、すべて原文の日本語によって決定されます。そこで、前編集が重要になってくるのです。

ここでは、日本語の特性や英語との違いを考え、前編集するうえで必要になる知識について説明します。

日本語の大きな特徴は「**曖昧性**」、「**婉曲表現**」にあります。A にもとれる、B にもとれるといった、どっちつかずの表現や、目的をあえて明確にしない間接的な言い方にこの特徴が表れています。

よく政治家が使う「遺憾に思う」や「善処する」とい

1　Google 翻訳は 2016 年従来の統計機械翻訳に替えニューラル機械翻訳技術を導入。人間の脳神経細胞の仕組みを模倣したもので、データから複雑なパターンを学習、ニュアンスを考慮した翻訳を可能にした。

2　DeepL 翻訳は、2017 年 8 月にサービスを開始したニューラル機械翻訳サービスで、ドイツのケルンに本拠地を置く DeepL GmbH が開発。

3　情報通信研究機構の先進的音声翻訳研究開発推進センターは、ニューラル機械翻訳の音声翻訳アプリ『VoiceTra』を 2020 年に公開。テキスト翻訳に特化した『TexTra』も公開。

うのは、いったいどういう意味を含んでいるのでしょうか。本当にこういった言葉は通訳・翻訳泣かせです。

【例】遺憾に思う

（1）期待に応じられなくてたいへん申し訳なく思う、

（2）強く反対する、

（3）とりあえず反対を唱えておく

【答え】場合によって、(1)、(2)、(3) になります。

【例】善処する

(1) 早速対策を立てて、うまく処理する、

(2) とりあえず処理する、

(3) ホントはやる気がない、逃げ口上

【答え】本来は (1) の意味ですが、多分 (2) や (3) でしょう。

「A 社の業績は、現在下降傾向にあると言われている」といった表現もよく使われます。この表現は、「A 社の業績は、下降している」と同じ意味です。しかし、「傾向」という**婉曲表現**と、「と言われている」という**間接表現**によって断言を避けた文章になっています。

これらの曖昧性、婉曲表現、間接表現は、ときには文章を柔らかくするオブラートの役目を果たすこともあります。**露骨な言葉を好まない日本人の文化からきた表現方法**でしょう。しかし、翻訳をするときは、これらの表現が障壁になります。

「日本語と英語の違い」とは

日本語と英語の間には、異なる文化による「**発想の違い**」が存在します。これが言語表現に大きく関係し、翻訳をするときの障害になります。この 2 つの言語間に横たわる違いを理解しないと、「とりあえず英語になっていて、なんとなく意味はわかるが、英語的表現ではない」といった程度の翻訳結果しか出てきません。

次の表は、「日本語と英語の違い」を簡単にまとめたものです。

日　本　語	英　　語
曖昧で明言を避ける表現	直接的で具体的な表現
省略語 (主語などの省略) が多い	省略語は少なく、係り受けが明確
文が長くなりがち	比較的簡潔
無生物は主語になりにくい	無生物でも主語になる
結論 (話題の中心) が最後	結論 (話題の中心) が最初
難解な文章の場合、読み手の能力が問われる	難解な文章の場合、書き手の能力が問われる

日本語を書くときに、このような点を注意するだけで、正確で英語らしい表現にすることができます。ここでは次のポイントを覚えておいてください。

【ポイント】

・無生物も主語にする

・結論 (話題の中心) を最初にもってくる

例　題

原　文	そのホテルでは 500 人の客が宿泊できる。
翻訳文	500 visitors can stay in the hotel.

【前編集】無生物を主語にする

編集文	そのホテルは 500 人の客を収容できる。
翻訳文	The hotel can accommodate 500 visitors.

実践問題

無生物を主語に書き換えてみましょう。

原　文	（1）努力の結果、彼は成功した。
	（2）このレポートには、以下の内容が記載されています。
編集文	（1）彼の努力が彼を成功に導いた。
	（2）このレポートは以下の内容を記載しています。
翻訳文	（1）His effort led him to the success.
	（2）This report mentions the following contents.

翻訳対象となる日本文の種類

日→英が翻訳対象としている分野は、報告書、論文、マニュアルなどで使われる実用記述文です。**小説や詩歌など、いわゆる行間を読んだり文法的に不十分な文章は、日→英の翻訳に適していません。**

現在の日→英で翻訳できる文章の種類は、以下のとおりです。

	翻訳に適している	翻訳に適していない
分野	報告書などの実用記述文	小説・ドラマ・詩歌など、古文
種類	平叙文、命令文	会話文、感嘆文

3 クロスランゲージ社の前編集の考え方の誤謬

しかし、クロスランゲージ社の解説する前編集の考え方と提案される一連の作業は、その作業に費やす時間と労力を鑑みると、機械翻訳を使うメリットを半減させることになる。実際、前述の例文の原文と前編集した文を情報通信研究機構 NICT の『VoiceTra』と『Google 翻訳』に入力すると、結果は以下のようになる。前編集しない日本語を翻訳させても、前編集した場合と同じか

決して不自然ではない英文が出力する。英語らしい訳文を出すのに前編集は必要ないのだ。(★は前編集した文の『PC-Transer』による英訳、◇は原文の『VoiceTra』と『Google 翻訳』による英訳)

【原文】
そのホテルでは **500 人の客**が宿泊できる。
【翻訳文】
500 visitors can stay in the hotel.(『PC-Transer』)
◇ **The hotel** can accommodate five hundred guests.
(『VoiceTra』)
◇ **The hotel** can accommodate 500 guests.
(『Google 翻訳』)
原文では人が主語だが、翻訳文は無生物主語

【前編集した文】
そのホテルは 500 人の客を収容できる。
【翻訳文】
★ **The hotel** can accommodate 500 visitors.
(『PC-Transer』)
The hotel can accommodate five hundred guests.
(『VoiceTra』)
The hotel can accommodate 500 guests.
(『Google 翻訳』)
『VoiceTra』『Google 翻訳』の翻訳文は原文の訳文と同じ。

下記の例文でも (2) は「無生物主語の英訳文」を出力する。(1) も自然な英文だ。「無生物主語の英訳文」にする必要などないのだ。

(1)
【原文】
努力の結果、彼は成功した。
【翻訳文】
As a result of his efforts, he succeeded.
(『VoiceTra』&『Google 翻訳』)

【前編集した文】
彼の努力が彼を成功に導いた。
【翻訳文】
His effort led him to the success.
(『PC-Transer』: the は不要)
His efforts led him to success.
(『VoiceTra』&『Google 翻訳』)

(2)
【原文】
このレポートには、以下の内容が記載されています。
【翻訳文】
◇ This report contains the following:
(『VoiceTra』)
◇ This report includes the following:
(『Google 翻訳』)

【前編集した後に文】
このレポートは以下の内容を記載しています。
【翻訳文】
★ This report mentions the following contents.
(『PC-Transer』)
This report mentions the following:
(『VoiceTra』)
This report provides the following:
(『Google 翻訳』)

AI 翻訳は原文が膨大でも瞬時に翻訳する。このため、原文が膨大なら明らかに不自然な日本語がない限り、それぞれの文をいちいち前編集するようなことはしない。曖昧性の問題は、日本語文の作成者が「曖昧でない日本語を書くように心がければ良い」ことで、明快な日本語文を書けない人間は翻訳対象となるような文章の作成に携わる資格がない[4]。とにかく、最近の AI 翻訳は前編集に時間を浪費するには及ばない。前編集せずとも AI 翻訳は前編集した文の『PC-Transer』による翻訳より遙かに高品質の翻訳を出力するのだ。

それにクロスランゲージ社の示す前編集は、**「英語らしい英訳を知っている人間がその英訳になるように日本語を書き換えたもの」**である。わざわざ前編集してまで機械翻訳にかけずとも、その前編集者が最初から翻訳したら良い。前編集という作業プロセスを設ける意味がない

4　1980 年代後半までは大企業でも理工系の技術者が製品の取説を書いていたが、文章力がなく人が読んでも分かり辛い拙い日本語のことが多く、機械翻訳にも処理しやすいものではなかった。そこで前編集が行われていたが、1990 年代にはプロのライターが書き直したものが取説になったので、前編集は必要ない。特に、コンピュータの取説などは、バージョン・アップ版の 90%が同じ文章なので、旧版の訳文を踏襲する形で「違う説明箇所だけ」翻訳する。これに利用されたのが IBM 社の訳例参照ソフト『Translation Manager』やグローバルリンク社の『Power TranslatorTM』だ。

のだ。「無生物を主語とする英文」が英語らしい英文となるケースがあるからと言って、敢えて「無生物を主語とする不自然な日本語」に予め編集してから機械翻訳し、「無生物主語の英文」を出力するというのは本末転倒だ。

「翻訳文の対象となる日本語の文、翻訳の対象とならない日本語の文」と区分するのも、前世紀のルールベースの機械翻訳ソフトの翻訳能力が AI 翻訳より格段に低い頃の考えだ。翻訳ソフト『PC-Transer』は現在も翻訳能力が 90 年代からさほど向上していないことが原因だろう。AI 翻訳は、（報告書などの）実用記述文だけでなく、小説・ドラマ・詩歌なども翻訳できるし、もちろん会話も訳せる。特に自動通訳を標榜する『VoiceTra』は会話に強い。「小腹が減った」という発話は原文にない主語 I を入れた “I'm a little hungry.”（『VoiceTra』）や “I'm feeling a little hungry.”（『Google 翻訳』）となる。小説やドラマで使われる言葉は言語的には実用記述文と変わるものではない。文芸作品のカテゴリーに属する現代の詩も問題なく訳せるものが多い。

現代詩

学校、うん、教室にいるとぽつんと、一つの島に一人ずつがいる気持ちになる、それがきれいな島ならいいけれど。（井戸川射子の第 1 詩集『する、されるユートピア』（2018）より）

School, yeah. When I'm in class, suddenly, I feel like there's one person on each island. I hope it's a beautiful island.（『VoiceTra』）

When I'm at school, yes, in the classroom, I feel like I'm on one island, and I hope it's a beautiful island.（『Google 翻訳』）

どうしても会えない気がして、走って、走って、そういえば誰に会いたかったんだっけと気づいて、それでもうとまり方がわからない。（最果タヒ『天国手前』より）

I just can't seem to see you. Run, run, Come to think of it, who did I want to see? I realized.

But I still don't know how to stop.（『VoiceTra』）

I felt like I couldn't see him, so I ran, I ran, Come to think of it, I realized who I wanted to meet.Still, I don't know how to stop it.（『Google 翻訳』）

はだしの少女は 髪に紅い野薔薇を挿し 夕日の坂を降りて来る。

石だたみの上に 少女の足は白くやわらかい。

夕餉の水を汲みに 彼女は城外の流れまでゆくのだ。

しずかな光のきらめく水をすくって 彼女はしばらく地平線の入日に見入る。

果てしもない緑の海の彼方に 彼女の幸福が消えてゆくように思う。

（田辺利宏「水汲み」より）「きけ わだつみのこえ―日本戦没学生の手記 (岩波文庫)」

The barefoot girl had a red wild rose in her hair. Coming down the slope at sunset.

The girl's feet are white and soft on the cobblestones.

To get water for dinner, she goes to the stream outside the castle.

Scooping up the sparkling water of quiet light. She stares at the sun setting on the horizon for a while. Beyond the endless green sea. Her happiness seems to fade.（『Google 翻訳』）

それでは、俳句＆短歌はどうだろう。俳句[5]や短歌[6]などはその独自の形式までそのまま英語に移せる訳ではないので、その形式が担う独特のニュアンスを醸し出すのは苦手かもしれない。だが、AI 翻訳させてみると、ほぼその形式に準じるニュアンスらしきものを何となく伝えられそうな英文にはなる。

俳句＆短歌

いいねと君が　言ったから　７月６日は　サラダ記念日（俵万智『サラダ記念日』）

You said you liked it, July sixth is Salad Day.（『VoiceTra』）

Because you said you like this taste, July 6th is Salad Anniversary.（『Google 翻訳』）

5　俳句は世界でもっとも短い詩。作り方には、5・7・5 という 17 拍の短いリズムの中に、季語を入れるというルールがある。

6　短歌は和歌の一種で、物事や出来事を 5・7・5・7・7 の 31 音で表現する古典詩。最初の 5 文字を「初句」と良い、次から順に「2 句」「3 句」「4 句」「結句」と言う。

柿食えば　鐘が鳴るなり　法隆寺（正岡子規）
When I eat a persimmon, the bell rings. Horyuji?（『VoiceTra』）
If you eat persimmons, the bell rings, Horyuji Temple.（『Google 翻訳』）

川を見る　バナナの皮は　手より落ち（高浜虚子）
Looking at the river, the banana peel falls from the hand.（『Google 翻訳』）[7]

「曖昧性」、「婉曲表現」

「遺憾である」は「思い通りでなく残念 / 残り惜しく思う」などの意味がある。しかし、実際の使われ方はそれとは違う。不祥事を犯した企業が「弊社としても遺憾です」とか「北朝鮮のミサイル発射は遺憾です」と首相がコメントするなど、日本人にとっても素直に納得しにくい。「弊社としてお詫びいたします」とか「北朝鮮のミサイル発射には強く反対する / 抗議する」と言った方が良い。AI 翻訳は「中国の活動は遺憾である」を "China's activities are regrettable."[8]（『VoiceTra』）や "China's activities are questionable."（『Google 翻訳』）と訳す。（regrettable は「残念な / 悔しい / 遺憾」questionable は「人に疑問を抱かせる」）

クロスランゲージ社の前編集の解説には下記の説明がある。

「A 社の業績は、現在下降傾向にあると言われている」といった表現もよく使われます。この表現は、「A 社の業績は、下降している」と同じ意味です。しかし、「傾向」という**婉曲表現**と、「と言われている」という**間接表現**によって断言を避けた文章になっています。・・・**露骨な言葉を好まない日本人の文化からきた表現方法**でしょう。しかし、翻訳をするときは、これらの表現が障壁になります。

しかし、「弊社の業績は、現在下降傾向にあると言われている」を AI 翻訳すると、Our company's performance is said to be on a downward trend now.（『VoiceTra』）、It is said that our business performance is currently on a downward trend.（『Google 翻訳』）となり、何ら障壁にはならない。日本語の文には婉曲表現、間接表現が多いかも知れないが、そのまま翻訳できるし、英訳文もまた不自然なものではない。

「発想の違い？」

解説では、日本語は「省略語（主語などの省略）が多い」、英語は「省略語は少なく、係り受けが明確」としている。だが、これはそれぞれの言語の文法的な特徴なのだ。決して発想の違いなどではない。英語では先行語と同一の語は繰り返さないが "Mary lost the book she bought yesterday" のように必ず「代名詞」という痕跡を残す。一方、日本語は先行語と同一の語は繰り返さないだけなく、欧州諸語における機能語である「代名詞[9]」も残さない。「真理子は [昨日買った] 本を失くした」のように省略するのが普通で、「真理子は [昨日**自分 / 彼女**が買った] 本を失くした」のように敢えて表すと不自然というかぎこちなくなる。特に、**彼女**がとするとこの**彼女**は真理子以外の女性を指すと理解されることもあるだろう。同一語を痕跡として代名詞で表示するか、痕跡を残さず削除するかどうかは、言語の文法的特性で、「発想の違い」として扱うのは、言語の特性に関する無知と呼ばざるを得ない。

★　★　★

筆者は 1997 年に『パソコン翻訳の世界』（講談社現代新書）を著わしているが、その第 3 章「機械翻訳と実務」の中で、前編集について次のように記している。

80 年代の初期モデルの段階では、機械翻訳システムの限られた処理能力を踏まえて、「制限文法」や「規格化日本語」によって原文を編集しようという提案があり、「全て単文に改める」「語彙は実質語も文法語も一義で使う」などの原則が数ページにわたり箇条書きされていました。これは極めて煩雑で神経を消耗する作業で、「自

7　分詞構文で主節の主語と一致していないにも関わらず主語が省略されているが、原文も破格なので英訳として許容可能ではないか。

8　Cambridge Dictionary : making you feel sad and sorry about something

9　日本語には欧州諸語における文法機能を担う「代名詞」はない。「彼 / 彼女 / 彼ら」は対象の指示ができ、一応「代名詞」とされるが、he/him/his、they/them/their などの格変化、性、数の一致などもなく、いわゆる文法機能語の「代名詞」とは違う。

動処理により人間の負担を減らす」という機械翻訳の理念に反しています。

日常の翻訳業務における前編集は、人間にも分かりにくい文を分かりやすく改める作業です。また、編集をせずにとにかく機械に通し、翻訳が失敗した部分だけ編集して、もう一度機械翻訳することもあります。前編集が合理的なものである限り、負担は大きくないでしょう。これまで実務で試みて有効であった「経験則」は、比較的簡単な指針の集合になっており、これを基準に前編集するのが現実的な対応です。

現在のように翻訳ソフトが個人に普及してくると、英日翻訳をしてみたいと思った場合、いちいち文書に手を加えるのはわずらわしいので、そのまま翻訳ソフトにかけるというのが実体です。特に、インターネットの英文は前編集しません。「内容をおおよそ把握する」という使い方が多い[10]ことのほかに、編集のためにはいったんファイル化する作業が必要になるからです。また、ダウンロード中に翻訳するソフト[11]では編集ができません。とにかく、翻訳ソフトを個人が使う場合には、企業が実務で行うような前編集はしないのです。ただし、日英翻訳の際などに原文がだらだら長過ぎる、修飾関係が曖昧で読んでもすぐに分からないような場合は平易な文章に改めます[12]。その方が訳文がはるかに良くなり、訳文をまともな表現に改める後編集の手間もかかりません。これは本来執筆者の文章力の問題であり、原文が分かりやすい文章になっていれば必要ないものです。編集というよりは「リライト」と言った方が正確です。

なお、同章では**翻訳支援システム**についても言及している。

原文を解析し、語彙、文法、修飾関係の誤りなどを指摘する**文章推敲支援システム**などを使って対話的に校正することもできます。このシステムでは、修飾関係が曖昧な場合、人間の判断を求めます。「壊れたラジオと時計[13]」は、ラジオだけではなく時計も壊れているのかどうか曖昧です。現在、並列や修飾などいろいろな関係の曖昧性について指摘するプログラム『**悪文推敲支援システム**』（NTT）が開発されています[14]。これは日本文を解析するものなので、こうした判断は前編集ではなく後編集の段階で求めることも可能です。とにかく優先度の高い関係で訳出するという設計にし、これが適切でないならば別の関係を選ぶというのが使いやすいでしょう。

ここで現在研究が進められている『**文章推敲支援システム**』について触れておきます。

（一般財団法人）日本特許情報機構では平成 26 年（2014 年）3 月に『産業日本語』を出版しているが、これはグローバルな特許情報の発信力の強化を目指し、高度な文書処理を効率よく低コストで実施するために、日本語の特許文書そのものの改善からのアプローチも重要と考えるものだ。日本特許情報機構は、平成 19 年度から「人に理解しやすく、コンピュータにも処理しやすい日本語」（産業日本語）を特許版・産業日本語委員会で検討を進め、『特許ライティングマニュアル』『特許ライティング支援システム』などを試作している。これらは特許文書を人間が作成する段階で「明快な文の作成を支援する」もので前編集とは全く違う。特許文に特化した「文章推敲支援システム」なのだ。

★　★　★

4 AI 翻訳は日本語らしい表現にも対応

英文の表現をそのまま生かした和訳をすると「英語の癖の強い垢抜けない日本語」になることもあるが、そうした場合は日本語として座りの良い表現に改めて和訳す

10　90 年代の日本のルールベースの翻訳ソフトの翻訳レベルは英日翻訳で 80 ～ 85%に留まる。一方、近似言語の仏英語間翻訳などは 95%前後になる。独英語間翻訳は 90%前後。やはり、近似言語で文法が 95%同じ日韓語間翻訳は 95%前後だった。文法や語彙がどれだけ近いかによって翻訳率が決まってくる。

11　日韓語間翻訳では文法や語彙だけでなく配列も同じため、リアルタイムの高度な翻訳が 90 年代でも行われていた。

12　日本の学校では明快な文章を書く訓練の授業がほとんどなく、これが分かりにくい文章の氾濫をもたらしている。特に理工系の技術者は文章力が高くないので、取説はプロによるリライトが必要だ。

13　[壊れた [ラジオと時計]] 、[[壊れたラジオ] と [時計]]

14　1997 年時点

る技法を教える必要がある。

たとえば、90 年代前半の翻訳ソフトの場合、英語では所有表現を用いた “I have three children.” を「私は [[三人] の [子供]] を持っている」と訳したが、それではぎこちないので「私には [三人][子供] がいる」のような存在表現に改める後編集が必要だった[15]。しかし、AI 翻訳では最初から「 [子供] が [三人] います」(『VoiceTra』)、「私には [3 人] の [子供] がいます」(『Google 翻訳』) を出力する。『VoiceTra』訳では会話において発話者として不要な「私には」も省略されている。ところが、“They have three children.” は「彼らには [[三人] の [子供]] がいます」(『VoiceTra』)、「彼らには [3 人] の [子供] がいます」(『Google 翻訳』) としっかり第三者 (彼ら) を訳す。かつては、「私には三人子供がいる」を “There are three children in me.” などに誤訳していたが、AI 翻訳では適切な『所有表現』“I have three children.” (『VoiceTra』 & 『Google 翻訳』) に変換する。また “Many students joined the club.” はかつての機械翻訳でも (AI 翻訳と同じく) 英文に沿って「多くの学生がそのクラブに入りました / 参加した」と訳せたが、“No students joined the club.” のような主語の名詞に否定数量詞 no を冠した場合は適切に訳せなかった。しかし、AI 翻訳では主語の否定数量詞 no を述語の否定に使って「そのクラブに入った学生はいませんでした」(『VoiceTra』)、「クラブには生徒は参加しませんでした」(『Google 翻訳』) と適切に訳す。90 年代の機械翻訳にはできなかったこうした日本語らしい表現への翻訳が、現在の AI 翻訳にはできるのだ。

5 AI 翻訳は文法による意味の違いにも対応

例えば、**仮定法では「事実と違う仮定」をする**ので、“He walked as if he were drunk.” では、実際には酔っていない。AI 翻訳では「彼は酔っ払ったように歩きました」(『VoiceTra』)、「彼はまるで酔っているかのように歩いた」(『Google 翻訳』) と訳す。これが**直説法だと「事実を仮定している」**。“He walked as if he was drunk.” では、実際に酔っている。AI 翻訳では「彼は酔ったように歩きました」(『VoiceTra』)、「彼はまるで酔っているかのように歩いた」(『Google 翻訳』) となる。『Google 翻訳』の訳文は「実際には酔っていない」仮定法の意味にも「実際に酔っている」直説法の意味にも解釈可能な訳になっている。

ちなみに、『VoiceTra』では (walked を worked に替え) “He worked as if he were drunk.”が原文となった英文の和訳は「彼は酔ったふりをして働きました」のように仮定法に相当する訳になる。ただ、直説法の “He worked as if he was drunk.” も同じ訳文になるので、仮定法と直説法の違いを適切に訳し分けているのではなさそうだ。なお、『Google 翻訳』では同箇所の訳に変化がなかった。AI 翻訳が依る大規模言語モデルに適切な素材がないのだろうか。

6 AI 翻訳には意訳はできない

AI 翻訳は高度化してはいるが基本的に直訳なのだ。意味を深く読み込んで、聞き手に話し手の意図することを理解させる翻訳 (**=意訳**) をするというのは人間でなければできない。しかし、学生に英文を訳させた場合、英文の表面的な和訳はどうにかできても、本当に意図された意味を理解できた訳になっているとは限らない。表面的というか上滑りな訳に留まることが多いのだ。その場合は、文脈に適合する意味を理解させ、それを和訳に反映できるようにする必要がある。

たとえば、identity という英語も日本人には難しい。心理学用語としては、英語をカタカナにしただけの「アイデンティティー」で良いが、一般読者にはそれでは腑に落ちる理解にはならない。辞書には、「正体」「身元」「個性」「自己同一性」などの訳語も載せられている。“She lost her identity as a teacher.” を学生は「彼女は教師としてのアイデンティティを失いました」や「教師としての自己を失った」と訳すかも知れない。それでも一応翻訳として成立するが、「教師として、いかに振舞えば良いか / どう学生に接すればいいか、分からなくなった」といった本人の思考に踏み込んだ意訳が状況的には学生にも腑に落ちるのではないだろうか。“He faced his identity crisis.[16]”を「彼はアイデンティティ

15 日本語の助数詞は「[[三冊の] 本] を読んだ」「[三冊] [本] を読んだ」「[本][三冊] を読んだ」「[本] を [三冊] 読んだ」のように、名詞句内部だけでなく、名詞句から遊離した位置にも現れる。

の危機に直面した」や「自己同一性の危機に瀕していた」と学生は訳すのだろうが、「果たして自分の生き方がこれで良かったのだろうかと思い悩んだ」の方が本人の心境を汲みやすい。また、“Settling in Japan was the equivalent of self-banishment: instant and eternal alienation.”を学生は「日本に住むことは自己追放すなわち瞬時に永遠に疎外されるに等しかった」と字義通りに訳すかもしれないが、それだけで英文の意味が本当に理解できたかどうかは分からない。「(欧米人が)日本に住むというのは、住んですぐに自分が社会からよそ者扱いされる思いをし、それがいつまでも続くことにもなりかねないことだった」といった意味であることは、日本語に訳して説明する必要がある。こうした意訳は、(翻訳者や英語教育者のような)文脈、状況、心理を深く読み取れる人間でなければできない。あくまで言語レベルの処理である AI 翻訳には到底望めない。

7 AI 翻訳には文化を踏まえた翻訳は期待できない

文脈だけでは分からない「文化に根ざした比喩的な意味」もある。たとえば、“John is a snake in the grass.”「ジョンは草むらの蛇です」では snake が「裏切り者」の意味で使われるが、「蛇」が裏切り者になる経緯は、旧約聖書[17]の創世記の「楽園追放」に記されている。アダムとイブは神に「その実は食べてはいけない」とされた禁断の木の実(りんご)を蛇に唆されて食べてしまう。その結果二人には知恵が芽生え、自らの裸の姿を恥じて恥部を葉っぱで隠すが、その姿を神に見られ戒めを破ったことが露見して、エデンの園を追放されることになった。この物語がキリスト教圏の人々の共有知識になっていることから、「蛇」=「裏切り者」という比喩的な意味になるのだ。授業では、そのことを説明しなければならないが、そうすることで訳読を通してキリスト教圏の文化教育を行なうことになる。

16 identity crisis=アイデンティティ(自己同一性)を喪失した状態。人が自己の役割、存在意義、目標などを見いだせず、混乱を生じたり、心理的な危機に陥ったりすることを伴う。

17 旧約聖書はもともとユダヤ教の聖典であるが、キリスト教だけでなく、イスラム教でも聖書コーランとは別に聖典として引き継がれている。この三宗教はいずれも同じ神(エホバ/ヤーヴェ/アラー)を信奉する一神教なのだ。

8 AI 翻訳には文脈情報を踏まえた翻訳はできない

AI 翻訳は文脈情報を踏まえて曖昧性(=修飾関係の潜在的な多義性)を解消することができない。例えば、下記の英文(および、仏語と西語への翻訳文)では、前置詞句 [with a telescope] が、①先行名詞 a girl を修飾するのか、②動詞 saw を修飾するのか、言語的には曖昧である。

(1) I saw a girl [with a telescope].
(2) Jai vu une fille [avec un telescope].
(フランス語)
(3) Vi a una chica [con un telescopio].
(スペイン語)

上記の例のように、欧州諸語間の翻訳では文成分が英語の原文と同じ配列で翻訳される。フランス語訳でもスペイン語訳でも、修飾関係が原文と同じ多義性を担ったままの訳文になるが、聞き手が状況を考えていずれか判断するので翻訳として問題ない。

ところが、英語から日本語への翻訳だと、この修飾関係の多義性を解消し、下記のような異なる訳文(a)、(b)いずれかにしなければならない。

(a)「女の子を[望遠鏡で]見ました」
(b)「[望遠鏡を持った]女の子を見ました」

AI 翻訳は文脈を読取ることができない。このため、言語的には修飾部の係り先が 2 つあり得るなど多義性がある場合には、どちらの修飾関係に翻訳しても正解になる確率は 50%になる。文脈に沿った形で適切に翻訳することは保証できないのだ。

一方、日英翻訳の場合は事情が違う。

下記の文は言語的には副詞の最近が (a) のように「開発された」にも (b) のように「開催された」にもかかり得る曖昧な文である。(b) の意味の場合にもこのように発話することは少なくない。状況によってどちらにも解釈できるのだ。

「最近アメリカで開発された車の展示会が開催された」

(a)「[最近アメリカで開発された] 車の展示会が開催された」
(b)「最近[アメリカで開発された] 車の展示会が開催された」
しかし、一般的には (a) のように近接した動詞を修飾するのが普通なので、(b) のように最近開催されたという意味であれば、(c) のように最近を開催されたの前に置くのが良い。
(c)「[アメリカで開発された] 車の展示会が最近開催された」

これは「修飾要素を被修飾要素に近い位置に配置する」という一般原則で、「**近接（配置）原則**」と呼ぶ。発話者はこの「近接原則」に沿った文を発出することが求められるのだ。それにより誤って解釈されることを避けることができる。原文に修飾関係の多義性がみられるなら、「近接原則」に沿った文に改めると、その原則に従う修飾関係の優先度が高くなる。人間にとっても正しい解釈がしやすいし、機械翻訳の入力として望ましい。もっとも、文章構成力が高ければもともと近接原則に沿った文を書くので、（前）編集まがいの作業は必要ない。

参考文献

『こうすれば使える機械翻訳』（成田一編著）バベルプレス 1994.4
『パソコン翻訳の世界』（成田一）講談社 1997.10
「機械翻訳はどこまで人間に迫れるか」（成田一編著）（『AI JAPAN』）白夜書房 2000.1
「特別講座：機械翻訳ことはじめ」（成田一）（『翻訳辞典 2002』）アルク 2001.11
「特許文の現代化と機械翻訳」（成田一）（『Japio 創立 20 周年記念誌』）（財）日本特許情報機構 2005.10
「特許文の多言語機械翻訳」（成田一）（『Japio YEAR BOOK 2006』）（財）日本特許情報機構 2006.11
「機械翻訳の歴史と今後の展望」（成田一）（『Japio YEAR BOOK 2007』）（財）日本特許情報機構 2007.11
「翻訳ソフトあれこれ」（成田一）（『私のおすすめパソコンソフト』）岩波書店 2002.8
『英語リフレッシュ講座』（成田一編著）大阪大学出版会 2008.4
「日本語編集の視座」（成田一）（『Japio YEAR BOOK 2008』）（財）日本特許情報機構 2008.11
『日本人に相応しい英語教育』松柏社 2013.8
『産業日本語』（平成 25 年度 特許版・産業日本語委員会報告書）（財）日本特許情報機構特許情報研究所 2014.3
「自動翻訳の高度化と英語教育」（『Japio YEAR BOOK 2019』）（財）日本特許情報機構 2019.11
「人間翻訳と自動翻訳」（『Japio YEAR BOOK 2020』）（財）日本特許情報機構 2020.11
「翻訳と通訳における機械翻訳の活用」（『Japio YEAR BOOK 2022』）（財）日本特許情報機構 2022.11
「言語の壁を超える AI 翻訳」（『Japio YEAR BOOK 2023』）（財）日本特許情報機構 2023.6