

機械翻訳と産業日本語

Machine Translation and Technical Japanese



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／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』（講談社現代新書）、『日本人に相応しい英語教育』（松柏社）、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）、共著『名詞』『現代の英文法 6』（研究社）、『私のおすすめパソコンソフト』（岩波書店）、『英語教育徹底リフレッシュ』（開拓社）ほか。論文・新聞（読売、朝日、日経など）・専門誌連載記事多数。英語教育総合学会会長。情報通信技術研究交流会運営委員。

1 はじめに

「人間翻訳と自動翻訳」『Japio YEAR BOOK 2020』で指摘したように、ここ数年の機械翻訳¹の翻訳レベルの向上が著しい。80年代後半からの文法規則をコアとする「規則型の翻訳ソフト」と90年代からの用例（対訳データ）ベースの「統計型の翻訳」²ソフトがあったが、この頃はいずれも実用レベルの入り口に達していなかった。2000年頃からは「ニューラルネット翻訳」（AI翻訳）³が研究されてきたが、2010年頃から統計翻訳にAI機能を複合する方式が採用され、1917年頃からは深層学習（Deep Learning）により翻訳能力が飛躍的

に向上した⁴。実用はかなり近づいた良質な規則型の翻訳ソフトを凌駕したと言える。

本稿では、こうした高度な翻訳能力を持つ現在の機械翻訳においては、『特許 writing マニュアル「産業日本語」』などのようなものは必要なくなっている現状を日英語の言語特性にも言及して解説し、最後に音声翻訳／通訳機の実用状況にも言及する。

2 産業日本語は時代遅れ

『翻訳の観点から日本語を再考する』「第七回 産業日本語を考える—機械翻訳の視点から」（WEB誌『Professional Translator』2021/04/07）において、小室誠一氏（バベル翻訳専門職大学院）は、

「世界的に見ると日本語はかなり異質の言語であり、欧米言語間の機械翻訳に比べて出力結果の品質は劣り、十分に活用されているとは言えません。その原因の一つが、日本語ライティングのスキル不足⁵と言えるかもしれません。日本語原稿が適切でなければ、当然、正しい訳文が出力されません。」（下線筆者）と述べている。

「日本語はかなり異質の言語であり、」「日本語原稿が適切でなければ、当然、正しい訳文が出力されません。」

- 1 従来「機械翻訳」と呼ばれてきたが、「自動翻訳」と呼ばれるようになって久しい。したがって、「自動翻訳」という用語を用いた方が良いのかもしれないが、学会でのテーマに「機械翻訳」が使われることが多く、本稿の引用シンポジウムも「第七回 産業日本語を考える—機械翻訳の視点から」『音声翻訳の社会展開の経過と機械翻訳の最前線に関して』であり、Japio YEAR BOOKの寄稿のジャンルも「機械翻訳技術の向上」なので、敢えて「機械翻訳」を使う。
- 2 大量の単語やフレーズごとの対訳データを解析し、その統計結果から適した訳し方を割り出す仕組み。
- 3 近年、「統計翻訳」に「ニューラルネット翻訳」を加えた複合方式が提案されている。ニューラルネット翻訳では、単語の意味だけでなく接頭辞や語幹、単語の位置なども考慮し、自然な文の流れを分析して翻訳する。複合方式では地の文の部分はニューラルネット翻訳によって翻訳し、専門用語部分は統計翻訳によって翻訳する。Google翻訳ではこの複合方式を英中翻訳に採用したが、日英翻訳の訳質の向上は複合方式化に起因すると思われる。

4 中堅大学の上位生の英日／日英翻訳能力を超えるレベルに達し、専門用語は翻訳者も及ばない。高校や大学では学生が翻訳ソフトで訳文を用意していることを前提にした授業が増えている。

5 文章（作成）力の不足ということ。

というのはその通りなのだが、「欧米言語間の機械翻訳に比べて出力結果の品質は劣り…その原因の一つが、日本語ライティングのスキル不足と言えるかもしれません。」⁶ というのはミスリーディングだ。

欧米言語間の機械翻訳に比べて訳質が低いのは「(英語やフランス語、ドイツ語など) 欧米諸語⁶ と日本語は世界の言語の中でも文法も語彙も最も異質な言語である」ことが原因だ。従来の規則型(ルールベース)の機械翻訳は「翻訳対象言語間の距離」⁷(特に文法装置の違い)が訳質に直接関係する。欧州諸語は文法や語彙が共通なものが多く、翻訳システムの構築がさほど困難ではない。逆に、両言語が異質であればあるほど、両者を結ぶ規則が複雑かつ多様で、システムの構築に膨大な知識と労力を要するのである。

これは外国語習得でも同じだ⁸。欧州連合 EU 諸国では中高生や大学生は母語以外に 2 言語を話せるし読み書きできる⁹。これは欧州諸語が統語操作の同じ文法装置や(音韻変化はあるが)語彙が共通なものが多いことが原因だ。一方、日本人は長年苦闘しても英語が習得できないが、それは英語が日本語と違いすぎるためのものであって、

6 北米は英語とフランス語、中南米はスペイン語とポルトガル語が公用語として使われているので、欧米諸語=欧州諸語として記載する。

7 英語からの言語的な距離を測ると欧州諸語は 1 だが、日本語や韓国語は 6 とされる。

8 「日本の英語教育」が期待される成果を生まなかったのには言語的な理由がある。第一の理由は、外国語の習得には「言語的な距離」ないし「言語差」が決定的に影響するということだ。国際英検 TOEFL の成績は、北欧諸語が毎年トップを占める。これは英語と文法が近く転用可能なためだ。一方、英語とかけ離れた日中韓語は下位に低迷してきた。一方、日本語能力検定は、韓国が断トツ一位で、漢字を使う台湾、中国がこれに次ぐ。近似言語に比べ異質な言語は習得期間が 6 ~ 9 倍になる。日本人が英語習得に苦しむのは言語差に起因するのだ。英語教育のあり方以前の問題だ。第二の理由は、特に口頭運用に関わるが、英語には数や時制の一致計算や無限移動操作を含む文法処理が含まれ、その脳内処理がハンディになる。人間の思考は脳の「作業記憶」における活動だが、リアルタイムの処理の時間と容量に制約がある。

9 EU の言語教育政策の指針 CEFR (ヨーロッパ共通参照枠) に母語と 2 外国語の習得が規定されている。その基盤をなすのが「複言語主義」。これは、言語を文化と一体化したものとして捉え、個人が複数の言語の学習を通して多様な文化体験を持つことを目標とする。実際は、英語が第一でフランス語ないしドイツ語やスペイン語など職業機会に有利な言語が選ばれる。

(英語運用力獲得の妨げになっていると長らく非難されてきた)「文法訳読」中心のかつての日本の英語教育が根本的に間違っていたのでもなければ、日本人が皆外国語音痴なわけでもないのである。「英語が話せない」のもリアルタイムで異質な文法操作を脳内で実行することが難しいためだ。欧州人は母語と同じ文法操作を脳内で実行するので、運用もスムーズなのだ。また、韓国人の俳優や歌手は日本語を流暢に話す。これは両言語の文法が 9 割 5 分同じ文法装置を持ち、漢字を背景とした語彙が共通なためだ¹⁰。日本人も韓国語ならば短期間で運用できるようになる。このことは筆者がかかり前から読売新聞¹¹や朝日新聞¹²のコラムや多くの雑誌でも指摘している。

90 年代初期頃までなら理工学部出身の製品開発技術者のライティングのスキル不足があって読みにくいマニュアルが多数あったことは確かだが、それ以降は製品説明などを読みやすく書くのはプロのライターになっている。それにライターでなくとも、文章を書くスキルは個人差が極めて大きく、機械翻訳の訳質が低い原因の一つに特に取り上げるのは実態を反映していない。

小室氏はご自身の認識から機械翻訳のための「産業日本語」¹³について下記のように論じている。

「産業日本語」のコンセプトで特徴的なのが、ライティングのプロセスに焦点をあてていることで、単純に、読

10 輸出入貿易への依存度が国内総生産の 80% を悠に超える韓国では、高校だけでなく最近では中学でも外国語を二つ履修することが義務付けられている。第一外国語はもちろん英語だが、第二外国語には日本語か中国語もしくは欧州語を選択する。反日教育が行われていた時期も、母語を活用できて習得しやすい日本語を選ぶ学生が圧倒的に多い。

11 『英語教育の現代的課題』「論点」(読売新聞)1997.11.29 - 日英語の言語差が世界で最も離れているので、習得が難しいことを指摘。

12 『英語の社内公用語 - 思考及ばず、情報格差も』「私の視点」(朝日新聞)2010.9.18 - 特に、WH 語の(節境界を越える)移動操作(WH-Movement)が、その操作を欠く日本語の母語話者にとって会話などリアルタイムでの運用上の難点であることを世界でも初めて指摘。

13 「産業日本語」: Japio (一般財団法人日本特許情報機構)による造語で、「産業・技術情報を人に理解しやすく、かつ、コンピュータ(機械)にも処理しやすく表現するための日本語」と定義されている。



みやすい日本語文章を書くだけでなく、目的によってどのように処理していくかを、概略次の通り規定している（下線筆者）。

「試みる日本語」：メモ的な文章で、思考のツールとしての日本語。

「伝える日本語」：読み手が効率よく間違いなく読み取れるようにする日本語。

「訳せる日本語」：外国語へ直訳できる日本語、焦点を文レベルに絞る。

「機械が訳せる日本語」：容易な後編集を伴って正しく機械翻訳できる日本語。

このマニュアル通りに原稿を作成していれば、翻訳会社や翻訳者が受け取った原稿を「訳せる日本語」に修正する必要はないのですが、日英翻訳業務でこの修正作業が多かれ少なかれ発生するのが実情とし、具体的に『特許 writing マニュアル「産業日本語」』に挙げられている7つのルール¹⁴を示す。どれも基本的な事項で、少なくともこれらのルールに従って作成すれば読みやすい文書、翻訳できる文書になるはずだ。

筆者は「産業日本語」というのは、歴史的な使命を終えたのではないかと考えている。現在の機械翻訳はGoogle翻訳など「ニューラル翻訳」¹⁵の登場によって格段の進歩を遂げ、もはや機械翻訳のための前編集といった根気のいる不自然な作業を人間が時間を割いて行う必要がない段階にきている。もちろん、原文と訳文を対照して訳の「抜け落ちや間違い」がないかを確認したり自然な表現に改めるといった（後編集）作業は必要だが、機械翻訳し易いように原文を意識的に書くとか前編

14 1. 短文にする～短くシンプルな文にする。2. 省略しない～隠れている要素がないか注意する。3. 理解しやすい構成にする～文の構造に注意する。4. 横並びの要素の表現を揃える～対等にならば意味に注意する。5. 読点を工夫する～係り受けや文の構造を明らかにする。6. 簡潔にする～シンプルな表現にする。7. 言い換える～誰にでも伝わる表現にする。

15 「ニューラル翻訳」は膨大な統計データベースとAI深層学習により訳質は飛躍的に向上したものの、ルールベースの機械翻訳と違い、それぞれの訳文を生むアルゴリズムは見えないものとなっているので、どういうところで誤訳や訳抜けが起こるか分からないのが難点だ。

集するという作業は思考と時間の浪費だ。現在の深層学習機能を備えた高度なニューラル翻訳を利用する際に、そういう前編集的な作業を介在させるのは、宝の持ち腐れにも等しい。高機能なニューラル翻訳はそういう作業が必要な代物ではないのだ。人間は翻訳しやすいかどうかを考えながら文章を書く必要はない。ほかの人にも分かる「明快な文章」を書くように心がければ、それで良い。それが文章を書く際の普通の思考プロセスだ。

翻訳方向：従来のルールベースの翻訳では、翻訳方向による品質の格差が大きい。英日翻訳は中堅大学の平均学生に匹敵し8割台の翻訳率だが、日英翻訳は6割台に留まる。（2000年以降の）統計翻訳をAI深層学習機能でサポートする（2017年頃からの）ニューラル翻訳では、英日は中堅大学の上位学生を越え9割台の翻訳率だが、日英でも8割台に迫るようになった。

3 日本語は構造保持言語ではない

産業日本語研究会シンポジウム『サイバースペースにおける産業日本語』（2021 / 03 / 05）の中のパネル討論において、「日本語は文脈依存的で省略が多く、英語への翻訳においてそれを補う必要があり、機械翻訳にはそれが難しい」との発言があった。

筆者はそれに対して次のような趣旨の指摘をした。

言語的には、英語は「文の基本的な主要成分を全てあらわさなければならない」**構造保持言語**であるため、繰り返される名詞を削除せずその痕跡として代名詞を残すのに対し、日本語は**非構造保持言語**のため、文脈的に分かっている場合には繰り返される名詞を削除し痕跡の代名詞を残さない。それは既知の文成分を省略しないと日本語らしい表現ではなくなるためだ。「明快な文章にする」以外に、機械翻訳の訳質改善に向けての編集方針としては、「日英翻訳では（英訳文で必要な成分である）日本語の省略成分を補う」という位が良い。それ以上に、すべて「短文にする」などの編集は煩わしく、文章構成の思考力を阻害する恐れもある。かつては翻訳困難だった「幾つか文を内部に含む多重な複文」でも今は機械翻訳できるのだ。

4 Translation Memory を使う翻訳

マニュアルなどの業務翻訳では、『SDL TRADOS (トラドス)』というソフトが使われる。これは、(たとえば、コンピューターなど短期間で新版の出る) 製品のマニュアルなどにおいて、旧版の訳文を再利用できるように、日本語と英語の対訳をデータベース化しておく翻訳支援ツールだ。マニュアルの翻訳では新旧版でほとんど同じ文章が大多数のため、業務ではこのツールを使用することが必須である。旧バージョンと比較および解析し、「どの程度の改訂率があるのか」を数値(ワード数や文字数)で確認することができ、一度翻訳した訳文を、次のバージョンアップ時に利用して、「訳文の統一」を図ることが可能だ。旧バージョンの翻訳時に、Translation Memory (対訳データ) が残る。

TRADOS を使用して解析作業を行う場合、厳密には、対訳データ内の文節と翻訳対象データ内の文節とのマッチングをする。特にコンピューターなど数年で新版の出る工業製品のマニュアルの説明文は 8 ~ 9 割が旧版と同じなので、日付などの変数部分を入れ替えて、新旧版で差異の見られる文章や語句だけを抽出して自動翻訳し、訳文を仕上げる。ほとんどの翻訳会社において Translation Memory が業務で使われている。

5 日英語の音声翻訳の現状

AC-Net¹⁶ のシンポジウム『音声翻訳の社会展開の経過と機械翻訳の最前線に関して』(2021/03/22) では、音声翻訳のプロセスについて説明があった。「語彙認識」について、英語が日本語より認識レベルが高くなるという報告があったが、理由の考察がなかったので、そうなるのは、(表意文字の言語である) 日本語には漢字を支えとした同音異義語が極めて多いのに対し、(表音文字の言語である) 英語には同音異義語があまりないことが大きな原因であると指摘した。(同じ音連鎖の語

が多ければ、その音連鎖がどの語かを決定することが難しい¹⁷ ということ。ただし、「分野限定によって」もしくは「ある程度までなら文脈からも」適切な語意選択ができる。) さらに、「日本語の音声認識と英語の音声認識」の違いについても、「英語では母音¹⁸ の数が多いだけでなく、強い強勢のない音節ではほとんどの短母音が曖昧な中舌母音 [ə] に変化するなど、弱音節における音韻変化が激しくどれもほぼ同じ音に聞こえるため、音声認識は英語の方が難しいのだ」とも指摘した。

6 通訳アプリと通訳機

音声認識と翻訳さらにその音声出力ができる「通訳機能」が充実した機器やアプリもここ数年で急激に増えている。スマホにダウンロードする無料の旅行会話アプリ「ボイストラ」¹⁹ (情報通信研究機構 NICT) は、訳したい言葉をスマホに吹き込むと、(京都府にある NICT の) スーパーコンピューターに送信され、瞬時に翻訳されて訳文が画面に送られ音声も流れる。この間 1 秒に満たないが、5G が使える地域だと、その短い間もなく瞬時というかほぼリアルタイムで会話が成立する。

自動通訳機ポケット (74 言語対応: 通訳 19,800 円/通訳+カメラ翻訳 29,800 円: 2017 年 12 月発売) やチータトーク (32 言語対応: 11,000 円: 2019 年 8 月発売) などを使えば、英語だけではなく多様な言語で日常的な会話だけでなく実務のやりとりもできる。病院や電鉄、家電店、百貨店など外国人の患者や客の多い職場でも本格的に導入されてきているが、外国に行く日本人もこれを購入しない空港で借用して、旅先での宿泊や買い物、観光に使うことが増えている。なお、ホテルや駅の大画面の掲示板に載せる複数の言語での表示は、基本的に人手による複数の翻訳を掲載するものだ。拡声器での自動翻訳も同じ。したがって、本質的に自動翻訳

17 「こうしん」は「更新、口唇、交信、行進、後進、高進、功臣、…」など多数の漢語に対応する。

18 日本語は 5 母音で、「ア」が一つなのに対し「ア」に近い音が英語は [ʌ] [æ] [ɑ] [ə] などに分かれ、さらに二重音や三重音もあるなど、20 前後の母音があるとされる。

19 Google 翻訳や「VoiceTra」(2010 年製品化は 2017 年頃に、Neural Network の導入によって、翻訳品質は格段に向上した。「VoiceTra」は基本的に会話場面の発話に対訳用例としているが、数十万例を超える用例を扱うため、広範囲な文に対応する。

16 “高度な情報通信”の英訳 Advanced Tele-Communications (AC) を名前に冠した AC-Net は、関西地区の情報通信分野の産官学の研究者・技術者交流の場として、1990 年 9 月に発足。多方面にわたる最新の研究動向・技術動向をテーマに数多くの講演会を開催。筆者も長年運営委員を務め、機械翻訳などをテーマとするシンポジウムを企画・司会している。



と言える代物ではない。

★★★

自動翻訳については、著書²⁰のほか日経新聞²¹や各種雑誌で記事を書いているが、日本の機械翻訳創成期とも言える80年代後半から90年代中盤まで日本電子工業振興協会「機械翻訳システム調査専門委員会」の学術委員を務め、機械翻訳開発企業（シャープ、日本IBM、NEC、リコー、凸版印刷ほか）やATR自動翻訳電話研究所などの研究機関から21回助成金を受けた。翻訳ソフトのスマホ導入に際しては、TBSラジオの実況インタビューも受けている。

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
- 「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

20『こうすれば使える機械翻訳』（編著）バベルプレス 1994.4、
『パソコン翻訳の世界』（単著）講談社 1997.10

21『「自動翻訳」膨らむ夢』「時論自論」（日本経済新聞）
1992.10



4

機械翻訳技術の向上

