

日本語編集の視座

大阪大学大学院教授
成田 一

PROFILE 英日対照構造論、機械翻訳、言語計画専攻。著書『名詞』（研究社）、『こうすれば使える機械翻訳』（パベルプレス）、『日本語の名詞修飾表現』（くろしお出版）、『パソコン翻訳の世界』（講談社）、『私のおすすめパソコンソフト』（岩波書店）、『英語リフレッシュ講座』（大阪大学出版会）のほか論文、新聞や専門誌、雑誌の記事多数。



1 はじめに

産業文書へのグローバルなアクセスが迫られている現代、国際共通語としての英語への翻訳が不可欠となっているが、人間による翻訳には膨大な時間と費用がかかる。このため、機械翻訳による処理が喫緊の課題となっているが、現状では翻訳対象となる文書の日本語が機械翻訳にとって問題のあるケースが多々あり、そうした問題を克服し状況を改善しようという試みの案がいろいろ提示されている。これには日本語の文書の側と機械翻訳システムの側にそれぞれ内在する特性を踏まえて、しっかりと見据えることが肝要である。

本稿では、特許文書における「日本語の明晰化」に関する①特許庁と②（機械翻訳の観点からの）東芝の提案を例に取り、機械による日英翻訳処理を念頭に置いた場合、「日本語の文章にどういった推敲・編集を施すことが適切なのか」ということを、日本語ならびに英語の言語的な特性を踏まえるだけでなく、機械翻訳の言語処理能力との関連で一般的に検討する。

2 日本語編集の視座

機械翻訳しやすいように日本語を推敲・編集すると言っても、それが「日本語の文法を恣意的に変更する」ようなものであってはいけないう、文書作成における「自

然な思考の展開に支障をきたす」ものであってもらいけない。さらに作業にあたる人間にとって「あまり負担になる煩雑な作業」であってもらいけない。文書作成をする人間と翻訳処理する機械側の機能のトレード関係を勘案することが大切だろう。

文書を作成する段階から「機械に処理しやすい文を意識して」作成作業を進めるとするのは、自然な思考の展開に支障をきたす可能性がある。したがって、そのような意識で文章を書くのは避けたい。ただし、人に良く分かる「明快な文章を書く」というのは、絶対に必要なことであるし、そうした意識で文章を書くのは決して自然な思考の展開を妨げるものでもない。これまで技術者が（関連技術分野の専門家以外の人にとっても）明快な文章を心がけて書くことが少なかったことから、「技術者の書いた文章は読み難い」とされることが多かった。（そういう文章は機械で翻訳しても良い訳文にはならない。）

ただし、これは技術者に限られるわけではない。日本の学校教育において（誰が読んでも）明快な文章を書く訓練がしっかり行われてこなかったことから、日本人の多くに見られる傾向であると言っても良い。もっとしっかり文章構成力を鍛える教育が求められるところだ。結局、「人間にとって読みやすい文章というのは、機械翻訳などにおいても解析しやすい文章になる」のである。

近年は説明書やパンフレットなどの商業的な文書の場合、専門のリライターが明快な文章に書き換えるケースが増えているが¹、現状は文章を書くことに慣れていな

い技術者が(弁理士が手を入れる前の特許文などを含む)産業文書を書くことの方がまだ多いだろう。そうした(リライトの処理の施されていない)文章は、「表現や修飾関係が明確でない」などの理由から、「意味が不鮮明だ」とか、「文章構成が貧弱だ」とされるものが少なくない。この場合、二つのレベルで改善作業を施さなければならないだろう。一つは「文章構成レベル」、もう一つは「言語レベル」であり、両者を区別することが肝要だ。文章の構成というのは「思考レベル」の問題であり、これと日本語の文の「言語レベル」の問題を混同してはならない。

本稿における「日本語の明晰化」という目標は、日本語文の「言語レベル」の問題を扱う。「文章の構成」は「思考レベル」の問題になるので扱わない。「文章の構成」が良いか否かは、人間が「文章全体の意味を理解しやすいかどうか」には関係するものの、機械翻訳の処理に直接関わる「日本語の明晰化」の埒外になる。したがって、本稿では、機械翻訳の処理しやすさ・しにくさを左右する「言語レベル」の問題に絞って日本語の処理の仕方を検討することとする。

3 日本語の規格化

機械翻訳しやすいように日本語を推敲・編集する際に配慮しなければならないことは、推敲段階での思考展開や編集作業の負担の面だけではない。機械翻訳しやすい文章になったとしても、それが日本語として文法的に間違っていたり不自然な表現になってはいけい。そうした言語面における問題がない編集であるべきだ。ところが、機械翻訳の研究・開発の歴史の中では、その当時(80年代中頃)の開発段階における言語処理の未熟さに摺り合わせるように、大学などの研究グループによって「日本語の規格化」案がまとめられたほか、(機械翻訳開発に携わる企業と大学関係者が参加する)日本電

¹ 英米の広告やマニュアルなどの英文は(特に「製造物責任(PL)法」対策もあり)誤解のないように平明に書かれている。

子工業振興協会の機械翻訳調査専門委員会において、「制限文法」案がまとめられた。いずれも日本語の文章を恣意的に変更したり文法的に制限する内容であった。細かい編集規則集だが、煩雑すぎて実務では使わないといつか使えない。そこに列挙された編集規則を適用すれば機械翻訳の正答率が上がったかもしれないが、中には日本語を歪めるようなものも少なくなかった。

驚くべきことに、機械翻訳とは関係なく、外国人への日本語教育という立場から、一種の学習文法として奇異な規則が提案されたことがある。日本語の文法では、丁寧体の「ます」は主文の述語に付加する。埋め込み文の述語には付加しない。ところが、国立国語研究所長で「簡約日本語」を提唱した野元菊雄氏は、「外国人にも学習しやすいように」という意図の下に、その日本語の規則の適正な適用対象を無視して、埋め込み文も主文も関係なく、「文の末尾には丁寧体の「ます」を付加する」ことを提案した。だが、「[[明日は雨が降ります]と思えます]」のように、埋め込み文にまで丁寧体の「ます」を付加するのは日本語として容認されないのだ。日本人なら決して使うことのない言語規則を勝手に作り出すこの提案には強い反対と批判がみられた。「自然な日本語の仕組みを歪めてはならない」ということだろう。それに、『丁寧体の「ます」は主文の述語にのみ付加する』ということは、外国人であっても、教えれば簡単に学習できる。野元案は学習者の能力を見くびった「おせっかいな提案」と言わざるを得ない。

他山の石ではないが、機械翻訳による処理を念頭に「日本語の規格化」を期する場合、日本語の文書の編集規則として、こうした非文法的なものを恣意的に設定する過ちを犯してはいけい、ということはいっしょに記録すべきだろう。

4 どんな前編集・後編集が必要か?

ここで「日本語の規格化」や「制限文法」において提案された編集案の内容を見てみよう。前編集とは「(間



違い)というわけではない」など装飾的な言い回しを「(間違い)ではない」に改めたり、「(操作)してみた」を「(操作)した」にすることで「てみた」が「て見た」に解析され operated and saw になるのを避けるなどの作業だ。こうした作業は手間がかかり、仕事でない限り、誰もやりたくない。幸い機械翻訳の解析能力が向上に伴い、これらの表現は正しく処理されている。「(嫌い)みたい」の「みたい」を本動詞に解析し、wants to watch itと訳すソフトもあるが、辞書の語彙数の増強などだけでなく、翻訳のコアになる言語的な解析の開発を継続して行っている企業のソフトは、編集なしで適切に翻訳できるようになっている。

本動詞ならば「見たい」と漢字を用い、助動詞ならば「みたい」と表記するというように、適切な「漢字仮名混じり」の文章ならば、正しく解析・翻訳するのだ。「漢字仮名混じり」の文章はこうした文法的な要素と(名詞、形容詞、副詞、本動詞などの)実質語を人間が区別するだけではなく、同音異義語についても漢字表記した時点で多義性が解消された表記となっている。このため、機械が知識処理することを回避できる入力となっているのである。そうした意味では、文章を易しく見せるためか、漢字を避け仮名表記にする²近年の傾向は、機械処理の観点からは望ましくないし、ひいては機械による翻訳文が用意しにくいという点で、グローバルなアクセスの機会を減少させるネガティブな効果をもたらすものである。

日英翻訳の場合、(明快な文章を書く授業が極めて不十分なためか)人間が読んでも原文が分かり難いことも少なくない。それを「普通に読んだら分かる表現に改める」というような編集は認めて良いだろう。これは編集マニュアルに沿っていちいち書き換えるのではなく、文章をしっかり推敲するプロセスそのものなのだ。

しかし、基本的には、「原文をそのまま処理する」のが機械翻訳の本筋だ。そうでないとコンピュータの俊足な処理のメリットが得られない。前編集が必要になるのは、「その文章が(悪文で)機械翻訳に適していない」か、「翻訳ソフトの精度が実用的な水準に達していない」

ということが原因だ。(英日翻訳が実用段階にあるのに対し、)現在でも日英翻訳の精度はやっと実用の入り口に達したに過ぎないが、それも原文の日本語文書がしっかりした日本語の文章になっている場合に限られる。人間が読んでも内容が読み込みにくいほど文章が乱れている場合には、機械による解析も全くうまく行かない。逆に、明瞭な文章であればどうにか実用的な翻訳になることが多いということだ。また、専門辞書が組み込まれていれば、それに対応する専門分野の文書でもある程度まで実用レベルの翻訳ができる段階にある。

5 特許文の翻訳

産業文書のカテゴリには入るのだろうが、その言語的特異性のために、極めて読みにくく機械による構造・意味解析が困難なのが特許文である。中でも「請求項」については、「これが日本語と言えるのか」と疑問を持たれるような複雑怪奇な文章もあり、(前処理以前の)文章作成の段階で根本的な改善が待たれる。そのような奇文を生む主因が「特許文の請求項は一文で書く」という慣習³とそれを実現するために、弁理士が用いる請求項作成手法である。

この慣習的な「請求項一文条項」を遵守しようとして、技術者の作成した説明文を基に、弁理士は、①いくつも文を内部に埋め込んだり、②いくつも名詞句を並列することにより、十行から二十行で一文となるような複雑な文を作り上げる手法を駆使する。だが、そうした文は文法関係、修飾関係が極めて捉えにくいものとなっており、機械による解析を寄せつけない代物となっていることが少なくない。

² 近年、標識などでも「チカン」や「ち漢」のように書かれるのを見かけるが、やはり漢字で「痴漢」と表記されてこそ、変質的な行為をする暴漢という意味合いが伝わるのであり、仮名表記では意味合いが感じられない。

³ 特許文の請求項の文は接続詞や接続助詞(英文ならば、カンマとか、ing形とか)などを手がかりに分割する。複文を切って行く処理を自動的にやる前処理プログラムも開発されているが、まだ試作段階の域を出ていない。しかし、そうした処理がうまくいった箇所については、特許文でもある程度は普通の文と同じように翻訳可能になる。

この弁理士の請求項作成手法によって作成された文書は「機械翻訳の対象外」とであると断ぜざるを得ない。その意味でこの弁理士が暗黙裡の内に使用する請求項作成手法はいわば「悪文の文法」と呼ぶのが妥当であり、こうした「悪文の文法」を捨てないことには、膨大な特許文を翻訳し外国からのアクセスに応えるという現代的な使命を果たすことは到底できない。これからは「特許文の請求項は一文では書かない」という原則を設け、単文を箇条書きにするようにして、複数の内容を記述するようにすれば、機械翻訳しても適訳が得られやすい。

ただし、その場合でも、訳文の精度を求めるならば、後編集は必要だ。これは翻訳システムが出力した訳文を、人間が原文に照らして訳語や修飾関係などの誤りを修正し、文体的にも洗練された訳文に仕上げる作業のことだが、文化的に誤解のない表現にするなど、翻訳文を読む人の理解力を念頭において、最終的な訳文に練り上げる。（編集作業を支援する環境として、欧米の翻訳者は、「翻訳者用ワークベンチ」というシステムを利用してきたが、現在日本で市販の翻訳ソフトには英和・和英辞典や用例集を搭載するものが少なくなく、その機能を担っている。）

6 現在の日本語前編集案

『産業日本語ワークショップ～産業情報基盤確立のための明晰な技術日本語である産業日本語～』（平成20年度：（財）日本特許情報機構）においては、「望ましい明細書に関する特許庁の取り組み」（特許庁）、「特許文書処理と明晰日本語—機械翻訳の観点から—」（東芝）という講演において、日本語の前編集に関する提案が示された。これは機械翻訳による処理を念頭に置いた最新の前編集案ということになる。以下では、その提案をいくつか取り上げ、本稿前半で論じた日本語編集の視座を踏まえて、その妥当性と改善点を検討したいと思う。

「望ましい明細書に関する特許庁の取り組み」（以下、特許庁案とする。）においては、「翻訳の観点からの留意

点」として、「用語の使用」、「文の記載」における留意点と「翻訳用日本語クレームの必要性について」ということを論じている。「用語の使用」における留意点では、1. 慣用語、2. 文化に基づく用語、3. 造語・複合語を扱っている。「文の記載」における留意点では、1. 主語の明確化、2. 指示代名詞の使用回避、3. 主語と述語の対応関係、4. 文の長さ、5. 助詞の使用を取り上げている。

「特許文書処理と明晰日本語—機械翻訳の観点から—」（以下、東芝案とする。）においては、翻訳結果の改善に効果のある規則として、下記の内容から成る「機械翻訳用明晰化規則」を定義している。1. 短文化：文の分割、2. 短文化：箇条書きの利用、3. 括弧書き（挿入句）の利用による基本構造の簡潔化、4. 並列表現に伴って省略された用言、格助詞の補完、5. 語句の関係の明確化、6. 動作の結果を表す「もの」の具体化、7. 意味の多様な「する」「なる」を含む表現の言い換え。

7 「単文を使う」という方針について

特許庁案では「一つの文に複数の主語や述語が混在すると、それらの対応関係について誤訳が生じやすい」という根拠を挙げて、なるべく「単文を使う」ことが提唱されているが、①単文を並べた文章というのは、（相互に因果関係がある場合などでは、）論理的な展開を追うのが結構難しいという面と、②（補文を埋め込む）複文であっても処理ができないわけではないという機械翻訳の機能面から、素直に首肯できる案⁴ではない。

「一つの文に複数の主語や述語が混在する」というのは、一文が複数の節（主節、従節）からなる場合を想定しているのだろうが、主節と従節が所定の論理的な関係で結ばれている場合には、これを切り離してそれぞれ単文で表すと、接続詞を介在させても、かなりぎこちない関係になってしまうことになりかねない。単文化し並列

⁴ 「単文化」というのは、かつての「規格日本語」や「制限文法」の提案の伝統を受け継ぐものである。



することにより、(複数の文からなりまとまった内容を表す)文章の意味的な結束性が弱くなったり、論理的な関係がつながりの悪いものになる危険性がある。文章中の文を全て単文に改めるということで主述関係が明確になったとしても、文と文との(論理・因果)関係が不分明になるようでは、読み手にとっても望ましくない文章ということになってしまうのだ。(特に、理由を表す節を主文から独立の単文に変えると、かなりぎこちない文になる。)

①主文に一つ文が埋め込まれる構造は、目的語や主語が文の場合には、当然生成されるものであり、これを単文化することは無意味に不自然な文に変える機械的な作業にすぎない。「 α が β に[冷蔵庫が故障した]ことを伝えた」を「 α が β に次のことを伝えた。[冷蔵庫が故障した]ことだ。」ないしは「冷蔵庫が故障した。そのことを α が β に伝えた。」と変えるのは、自然な思考プロセスに沿った言語化とは言いがたいし、本来の目的である機械翻訳の誤訳を減らすということにも寄与しない。現在の機械翻訳はこの程度の複文構造を誤訳するような低いレベルの言語処理能力ではないからだ。

8 曖昧な助詞の回避 (基本助詞⇒複合助詞)

基本的な助詞は複数の意味を持つものが多い。機械翻訳においては多義であることは誤訳の原因になる。特許庁案では、これを回避するために、意味が曖昧ではない助詞に変えることを望ましいとしているが、そうした助詞は基本的な助詞ではなく複合的な助詞である。たとえば、基本助詞「で」は「によって」(手段)、「において」(場所)に置き換えることになる。基本助詞「と」は(A and Bに相当する)『付加』の意味に限定し『随伴』の意味では複合助詞「とともに」を使うようにする。機械翻訳の助詞の処理もそれに対応する設定にするのが適当だろう。

9 慣用語句・表現の回避

特許庁案では、慣用語句や特定の国・地域の文化に根ざした語句・表現は、ほかの国・地域の人間にとって分かりづらかったり、誤解されるということが少なくないので、避けるべきであるということになっている。確かに、「目から鱗が落ちる」「逆手に取る」をそのまま翻訳しても意味は伝わらないだろう。

多国間コミュニケーションの場合は、地域文化に根ざした語句・表現は避けるのが基本である。このため、世界共通語になっている英語を使った国際コミュニケーションの場においてはその基本が守られる。いわゆる「国際英語」⁵における語句・表現は、地域的特徴を持たないものを規範とするべきだ。その意味では、英語を母語とする人であっても、相手の理解に配慮して英米の文化に根ざした慣用語句・表現の使用は控えるのがマナーということになる。

同様に、(異文化の人が読むことに配慮した意識を行うことのできない)「機械による翻訳」を前提とする産業翻訳の原文においては、訳文が直訳にしかならないことを念頭に、日本の文化に根ざした慣用語句・表現の使用は控えるという方針が望ましいだろう。ただし、そうした語彙や語句が機械翻訳辞書に登録されているものならば問題ない。

10 一般文書の用語と専門文書の用語

機械翻訳での用語処理については、①標準設定の一般辞書による処理か、②専門辞書による処理かを考慮しなければならない。

特許庁は「自然状態では」が“in the natural state”などに訳されることがあるという理由で、「自然状態で

⁵「国際英語」は(インド英語、フィリピン英語などの地域の母語の影響を受けた)「地域英語」とは区別されるべきもので、英語ネイティブも非ネイティブも使うコミュニケーションを円滑にするための中立的な言語としての性格を持たせたい。

は]を「無負荷状態では」に改めることを提案している。しかし、そうした変更が本当に必要なのだろうか。標準設定の機械翻訳では「自然状態では」が“in the natural state”などに訳されるが、これは一般文書への対応としては適切なものだ。したがって、特定の専門分野で「自然状態では」が「無負荷状態では」の意味で使われる慣習があったとすれば、一般辞書設定の機械翻訳を通す文書では「無負荷状態では」と記述しなければならないだろう。その点では特許庁の提案は適切である。しかし、専門辞書を優先する設定にした上での機械翻訳ならば、「自然状態では」という記述で構わない。当該の専門分野において「自然状態では」が「無負荷状態では」の意味で使われる慣習があることに対応した辞書を備えている場合には、「自然状態では」が「無負荷状態では」に自動的に変換（解釈）されることになるからだ。

ほかにも、特許庁は「異音」が言語学の用語“allophone”と訳されることがあることから「がたがた音」と記すことを提案しているが、これにも疑義がある。言語学の一分野である音韻論ないし音声学の専門辞書において「異音」が“allophone”と訳されるからと言って、そうした分野以外の文書において原文の「異音」を別の言葉に置き換える必要は全くない。それは機械や装置の「異音」というのが“queer sound”か“abnormal sound”のように訳されるのが標準であり、それが機械翻訳の一般辞書の設定としては妥当であるからだ。むしろ、提案されるように、「がたがた音」と記すことの方が問題だ。「がたがた音」は擬音語であり、これについては“rattling sound”と訳すことは可能だが、日本語と比べ擬音語が格段に少ない英語に翻訳するときには、対応する語彙や表現が欠けていることも十分あり得る。したがって、一般に機械翻訳を通す文書では擬音語（特に英語にはほとんどない擬態語）を含まない方針にするのが良い。

このように、①標準設定の一般辞書による処理か②専門辞書による処理かを考慮した場合には、専門分野の慣用的な用語や表現を無理に変更しなくても済むケースが多々あると思われる。

11 主語の明確化について

特許庁は、「日本語では主語を含まない表現が慣用的に多く用いられているため、…」と記しているが、これは誤解を招く記述だ。日本語において主語を含まない文が多いのは確かだが、主語が欠落するにはちゃんと言語的な条件があり、決して「慣用的に」といった（言語規則とは関わらない）変則的な理由で欠落するわけではない。（むしろ、主語の名詞が削除されるべき言語的な条件下において、削除しないで残すと不自然な表現になる。）換言すると、「日本語では主語を明示しない曖昧性を好む」文化的な傾向があると誤解されている理由で主語が欠落するのではないということだ。したがって、「慣用的に」という記述ではなく、日本語の言語的な特徴として、「先行名詞と同じ名詞は削除する」と記述しなくてはならないのである（英語では「先行名詞と同じ名詞を代名詞化する」、削除はしない）。

そうした主語を持たない文を人間が読む場合に、削除された主語が何を指すのか不明であることはない。同じものを繰り返すことを「情報的に冗長」とであると判断するからこそ、日本語では削除し、英語では代名詞という影武者に変えるのである。日本語では情報的な冗長性を排した表現となっているのに対し、英語では文脈的に分かっている情報を「文構造を維持するための成分」として代名詞の形で明示するのだが、これは余剰的な情報提示である。一般人の思い込みとは逆に、「日本語は簡潔な言語で英語は冗長な言語」なのである。

ただし、「外国語への翻訳という観点からは、明細書等において主語を明示した日本語を用いることが好ましい」という提案は、「機械翻訳による」外国語への翻訳ということ在意図しているのであれば適切である。英語は文の主要成分を（代名詞として）残すことにより「文構造を維持する言語」なので、日本語を英語に翻訳する場合、「削除された名詞を復元する」必要がある。（将来的にもその機能が備わる見込みは薄いだが、）現段階では、



先行名詞が削除された名詞と同一文中にない場合、そうした復元ができる「文脈処理機能」が機械翻訳に備わっていない。このことから、翻訳ソフトで処理する文書においては、(日本語的には何を表すかが分かっている要素を繰り返す冗長性があるものの、) 同一名詞を削除しないで明示するのが現実的だということになる。

「太郎が家に戻った。部屋に入ると、テレビを見た」のような先行文に照応すべき名詞がある場合には、「部屋に入ると、太郎がテレビを見た」に復元することはできない。一文を越えた照応関係の同定機能がないのである。これに対し、先行名詞が削除された名詞と同一文中にある場合は、削除操作の逆の(統語的もしくは解釈的)操作を行う言語規則を設定することができるので、翻訳ソフトがそうした規則を組み込んでいる限り、削除された名詞の復元ができる。「太郎は部屋に入ると、テレビを見た」を、「(太郎は) 部屋に入ると、太郎はテレビを見た」に変換することは、翻訳ソフトによる言語処理として可能であるし、英語に翻訳するときに、次のような代名詞を含む文に変換することも可能である。“After Taro entered the room, he watched TV.” でも良いが、“After he entered the room, Taro watched TV.”の方が自然な英文である。

同一名詞の削除と復元は主語だけではなく、目的語などでも問題になることだが、目的語の場合でも主語の場合と同じ扱いが妥当だろう。「構造維持的な言語」である英語においては、文を構成する成分をできるだけ表現上も具現させる。主語が必要だし、目的語も一部の動詞を除いて省略できない。そのために仮主語とか仮目的語といった形式的な要素を導入したり、既出の名詞を省くのではなく代名詞によって形態的に表現する。一方、「構造維持的な言語」ではない日本語では、そうした要素の残存は必要なく、情報的に分かっているものはほとんど省略する。文法上の照応を支える代名詞といった要素もない。このため、日英翻訳にあたっては、日本語の原文において欠如する要素を復元し、文の構造を維持する成分を表す必要がある。英語の文構造を構成する主要な要素を「復元する仕組み」が必要だということだ。そ

ういう作業を本格的に遂行するには、やはり人工知能的な文脈処理機能が要る。

このように日英翻訳においては、文構造を構成する主要な要素を「復元する仕組み」が必要なのだが、翻訳ソフトによっては、言語的な構造の選択などによって、「復元作業を回避する仕組み」も工夫されている。たとえば、「最初に議論すべき問題」は連体修飾節になっているが、節内には主語がない。主語を復元しなければいけないと言っても、必要な情報が節内にあるわけではないので、システム的には復元もできない。このような「主語を欠く」ケースにどう対応しているかだが、A problem to discuss firstと訳した場合のように「不定詞句」を使えば主語を表さなくて済む。The problem which should be discussed firstのように「受身」にしても復元が回避できる。さらにThe problem that you should argue in the first placeのように「総称的代名詞」のyouを入れても良い。ただ、残念ながら、The problem that it should discuss firstのように、主語が欠けている場合に「itを入れる」という設定になっているソフトもある。かなり木目細かく対応しているものとそうでないものがあるのだ。ただし、こうした言語的な対応に失敗した例としては、The problem which should argue (at) firstのように、「主語を補わない」というか論理的におかしい直訳も見られる。

12 曖昧性の処理－言語レベルの工夫

日英語のような言語差の大きい言語間の翻訳には、大きなハードルがある。ある文がどういう構造的なまとまりを成しているかを解析するだけでなく修飾関係も判定しないと翻訳できないことが多いのだが、それには文脈情報が欠かせない。例えば、I saw a girl [with a telescope].は「私は女の子を[望遠鏡で]見た」か「私は[望遠鏡を持った]女の子を見た」と訳し分けなければならないが、これにはwith a telescopeの修飾先を決定しなければならぬ。一方、英仏翻訳ならばJ'ai vu

une fille [avec un telescope].のように修飾関係が原文と同じく曖昧なままでも翻訳が成立する。日韓語や英仏語のような言語的に近似した言語間においては、修飾関係の判定をしなくても、構造的なまとまりを解析するだけで翻訳が可能なのである。このため、9割5分ほどの精度になっている。

こうした修飾関係の曖昧性を本格的に解消するには文脈情報に基づく知識処理が不可欠のだが、機械翻訳ではそうした知識処理機能が備わっていない。このため、修飾関係の曖昧な場合が3割程とすれば、確率的には1割5分の間違いを犯すことになってしまう。しかし、現実にはキーワードを始めとする言語的な手掛かりや統計的な傾向に基づく設定、言語的な条件による決定も可能だ。

工学的には修飾語に近い名詞を被修飾語を選ぶ「近接原則」を設定することがある。それで良さそうな「可愛い女性の靴」でも、「可愛い靴」の解釈の可能性は「可愛い女性」の解釈の可能性と変わらないだろう。一方「大きな庭の石」では「大きな石」の解釈のほうが「大きな庭」の解釈より可能性が高いと考えられる。単純な近接原則は統計的に65%ほどの正解にしかならないのだ。

一方、言語的な条件で修飾関係を決定することもできる。身体の一部は誰もが持つものであり、その（弁別的な）特徴を記述しなければ言語表現にならない。たとえば「瞳の女性」では意味がない。このため「きれいな瞳の女性」のような表現においては直前の形容詞「きれいな」は「女性」ではなく「瞳」を修飾すると決定できるのだ。「瞳」の代わりに「髪」「足」「うなじ」などが現れても同じ解釈になる。「身体の一部」という意味特性を持つ名詞の前の形容詞が直後の名詞を修飾する設定にすればこうした関係は処理できる。

分野によっては所定の修飾関係が優位なため、構造条件に応じて、現実には生じる確率の高い修飾関係を選ぶように設定すれば、正訳率は高くなる。たとえば、不定詞句は直前の名詞を修飾する形容詞用法と動詞を修飾する副詞用法があるが、一般の文章では副詞用法が多いので、そう解析するように設定する。さらに、副詞用法といっ

ても普通は「目的」を表すのに対し、技術文や自然科学、化学論文などでは「結果」を表すことがかなり多いので、「目的」を標準としても、(Water crystallizes to form snow. 「水は結晶化して雪になる」などに対応するように)「～して、～する/になる」といった「結果」としての訳の設定ができるようにする。(The hydrogen and oxygen were combined in the container producing water. 「容器内で水素と酸素が結合して水ができた」などのように) 分詞もほとんど「結果」になる。だから、分野を限定した機能として、そう設定すれば良いのだ。

13 定型的な表現の翻訳

特許文のように、(過剰使用の嫌いのある「前記の」など) 所定の定型的な表現が多用される場合には、その表現に対応する処理をする自動前処理ないし編集プログラムもしくは専用文法を組み込んだ翻訳ソフトも販売されているが、それによる言語処理は自動処理なので人間にとって負担はない。

産業翻訳一般について言えば、文書の構造を認識して、専用文法によって処理する可能性もある。たとえば、タイトルの位置に来ているものは、主語が抜けるとか、普通の表現とは違った形になっているが、文書の構造を認識して、タイトルの表現は本文と違う翻訳をするように設定する。「タイトル用文法」に切り替えて翻訳するのだ。また、タイトル文を普通文に内部的に変えるようなシステムもある。

「用例パターン」を用いる翻訳も有効だろう。大手電気通信系企業が翻訳システムの初期モデルを開発し終えた90年代前半には、「規則に基づいた従来の翻訳は限界にきた」として用例や統計をベースとした翻訳を探る動きが出た。当時、これは言語構造に関する専門的な知識が乏しいことに起因する対応だった面もあるが、言語的な知識の移植や言語構造の研究によって、翻訳ソフトの言語処理能力が向上して以降も、用例パターンを用いる翻訳機能(「用例翻訳」)を充実させて随所に「こなれた



翻訳」を入れる工夫も増えている。

用例翻訳というのは、ある言語における所定の表現を目標言語の表現に翻訳するために、双方の表現を翻訳対として設定し、入力と一致する表現をこの翻訳対のリストの中から検索して目標言語の表現に置き換えるものだ。この際、用例として設定されるのは単語列だけではなく可変な要素を含んで構わないが、通常これを変数で表す。

用例翻訳として市販ソフトに組み込まれているものは、対訳データの中から「典型的な表現を選択して用例として登録」している。用例を具体的な単語列として登録するだけではなく、可変部分に名詞句とか形容詞などといった文法範疇や意味素性まで付与することも可能だが、そうすると限りなく文法記述に近づく。用例を表現パターンとしてそこまで定式化したものは、用例翻訳というより用例パターン翻訳の方が適切かもしれない。用例パターンの記述には、「語彙列の指定がメインで変数部分にだけ文法カテゴリーを使う」ものから、「語彙指定が一部でほかは文法カテゴリーによる構造記述」という形式のものがある。本来パターン翻訳の記述は前者の形式をとるが、後者の記述になると限りなく文法ルールの構造記述に近づく。用例パターンには、汎用的なものと同分野特有のものがあり、タイトル表現など文書構造に依存するものもある。このため、キーワード照合などによって文書の分野を同定し、それに対応するパターン群を組み込んで翻訳する設定が良いだろう。類似例の検索は、対訳用例中の可変部分の指定にヴァリエーションがみられる。定型的な文の数字や日付などを自動的に置換して翻訳する「ファジーマッチ」のほか、可変部分の意味特性を指定して、We would be happy to send you < \$書類 > in return.を「折り返し<\$書類>をご返送します」に訳したり、名詞を変数にし、「[N₁]で[N₂]を持っている人が増えた」を [N₁] who has [N₂] has increased. に対応させ、「若者で携帯電話を持っている人が増えた」を Young person who has the cellular phone has increased.に訳すなどの方式がみられる。

14 マニュアルの翻訳

産業翻訳の一分野として「マニュアルの翻訳」があるが、これには「メモリー翻訳」が欠かせない。メモリー翻訳ソフト（トラドス、トランジット、SDL）が日本の翻訳の現場に本格的に進出してきたのは98年以降のことだ。しかし、それ以前（90年代前半）から、特定分野や個別企業における定まった用語や表現を使った翻訳の蓄積を生かすような機械による翻訳技術はないか、という要望はクライアント側にも翻訳会社側にもあった。これは用例翻訳にも反映されているが、「人間が翻訳した文章を対訳データベース化して再利用する」メモリー翻訳もほぼ同じ考え方に立つ。

メモリー翻訳は基本的に対訳データを使って「訳文に置き換える」操作を実行するにすぎない。その意味で構造解析と訳文生成を文法プログラムに基づいて実行する機械翻訳とは全く異質な操作だ。用例パターン翻訳が「典型的な表現を選択してパターン登録」するのに対し、メモリー翻訳は既翻訳の「生の対訳データ」をベースとする点で違うが、基本的に「単語列の検索と置き換え」という処理については同じ仕組みだ。

メモリー翻訳はマニュアルなど「変更箇所が少なく旧版の翻訳をそのまま利用可能な」文書に限り、既翻訳データを有効に利用するプログラムだ。したがって、翻訳データがない場合には使えない。翻訳データは翻訳を依頼する企業が著作権を持つので、その許可が得られる企業活動の中でしか使えないのだ。このため、業務使用に限定されることになるが、既に文書の現地化（ローライゼーション）を業務とする翻訳会社ではほとんどがメモリー翻訳を導入している。

では、メモリー翻訳にはどのようなメリットがあるのだろうか。機械翻訳と同じように、「訳語が統一され」「訳し漏れがない」などのほかに、訳文や訳語の共有により大量の翻訳をグループで担当したり訳文をみんなで検討することもできる。また、模範となる翻訳を新人が手本とするなど教育にも役立つという指摘もある。

さらに、構造解析を行うわけではないので、かなり高速に訳文への置き換え処理が実行される。このため、新たに翻訳しなければならない分量を簡単に算定でき、翻訳請負業務が効率的に契約できるのだ。しかし、日常的に使いこなすには1年ほどかかると報告されるほど習熟の難しいソフトなので、翻訳会社などにおいて集団的に使用するのに適していると言われる。

(本稿著者執筆の) 関連文献 (1997以降のみ)

- 『パソコン翻訳の世界』講談社1997.10
- 「翻訳ソフトの実力評価」(『情報処理学会研究報告』98-NL-125, pp.123-130情報処理学会[自然言語処理研究会]1998.5
- 「言語学フロンティア」月刊『言語』Vol.26, pp.114-119大修館書店1997.4
- 「翻訳ソフトの実力評価」(『情報処理学会研究報告』98-NL-125, pp.123-130情報処理学会[自然言語処理研究会]1998.5
- 「機械翻訳はどこまで人間に迫れるか」『AI JAPAN』pp.43-59白夜書房2000.1
- 「英日・日英機械翻訳の実力」『第6回年次大会発表論文集』pp.51-54言語処理学会2000.3
- 「機械翻訳の限界に挑む」『CAT』pp.33-35アルク2000.7
- 「グローバル・コミュニケーション・ツールとしての翻訳ソフト」『日経WebCOMPANY』pp.57-58日経事業出版社2000.10
- 「人間に迫る機械翻訳-人間との共同作業-」『IJET-2000』pp.86-110(第11回英日・日英翻訳国際会議プロシーディング)2001.6
- 「特別講座・機械翻訳ことはじめ」『翻訳辞典2002』pp.162-169アルク2001.11
- 「機械翻訳はなぜ見放されたのか?」(機械翻訳の薦め—その1)『Translators'』(84号) pp.10-11日本翻訳協会2002.3
- 「今こそ機械翻訳を見直そう」(機械翻訳の薦め—その2)『Translators'』(85号) pp.8-9日本翻訳協会2002.8
- 「MTの現状と今後[国内篇]」『eとらんず』pp.26-30バベル・プレス2002.8
- 「翻訳ソフトあれこれ」(『私のおすすめパソコンソフト』) pp.143-156岩波書店2002.8
- 「英語嫌いのための 最新・翻訳ソフト活用術」(企画編著)『SPA!』pp.69-81扶桑社2002.10
- 「今こそ機械翻訳を見直そう」(機械翻訳の薦め—その3)『Translators'』(86号) pp.12-13日本翻訳協会2002.12
- 「ブリッジ方式による多言語翻訳」(機械翻訳の薦め—その4)『Translators'』(87号) pp.12-13日本翻訳協会2003.3
- 「特許文の改善に向けて —特許文書の問題点—」『eとらんず』pp.62-65バベル・プレス2003.8
- 『Translators'』(94号) pp.25-27日本翻訳協会2005.5
- 「特許文の現代化と機械翻訳」(『Japio創立20周年記念誌 特許情報活用の時代の検索と機械翻訳技術』) pp.194-201((財)日本特許情報機構)2005.10
- 「特許文の多言語機械翻訳」(『Japio 2006 YEAR BOOK 特許情報活用の時代の検索と機械翻訳技術』) pp.210-221((財)日本特許情報機構)2006.11
- 「機械翻訳の歴史と今後の展望」(『Japio 2007 YEAR BOOK 特許情報活用の時代の検索と機械翻訳技術』) pp.214-221((財)日本特許情報機構)2007.11