

機械翻訳の歴史と今後の展望

どこまで人間の翻訳に近づけるか？

大阪大学大学院教授

成田 一

✉ narita@lang.osaka-u.ac.jp

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）のほか論文、新聞や専門誌、雑誌の記事多数。

機械翻訳の今後の展開を予想するには、これまでの研究と開発がどうなっていたかを知ることが必要だ。そこで機械翻訳の歴史を振り返るところから始めよう。

1

第一次ブームと第二次ブーム

日本において機械翻訳はこれまで2度ブームを迎えた。最初は90年前後に電気通信系主要企業が第一世代の実用システムをほぼ製作しワークステーション（WS）と一体で販売した頃で、ユーザは一部翻訳会社のほか一般企業だったが、これは92～93年頃で終わる。24ドットの漢字プリンターやスキャナーのセットで6～800万円の時代だ。93年頃にはWSとソフトで300万円まで下がったが個人には手が出ない。ソフトのみの販売も数社あったが90年前後は英日、日英それぞれ50万円で93年頃に20万円になった。パソコンとプリンターで100万円の頃だ。

第二次ブームは95年～97年頃にやってくるが、これは当時10万円まで翻訳ソフトが下がる中94年暮れに「コリヤ英和」が一万円を切って販売されたのがきっかけで、個人が一斉に購入に動いた。月に一万本と業界一年分の販売本数を一桁超え、他社もこれに追随した。このブームの背景にはバブル崩壊後のパソコン価格の急落により一般家庭にパソコンがかなり普及したこととウィンドウズ、インターネットの登場があることも見逃せない。

2

期待を裏切った翻訳精度

第一次ブームにおいては翻訳システムの実力は英日でも60%ほどの翻訳精度で70%あればトップクラス、日英はせいぜい50%台だった。広告の例文では一般の注意を引いたし新聞や雑誌も「夢の技術」と持ち上げたが、これはその例文用に予めチューニングを施して市販よりも訳語が改善するようにしたものだ。購入してみると期待はずれの翻訳でしかも読み取りや前編集に時間を割かれることが多く、結局放置される道を辿ることになった。

第二次ブームの頃に機械翻訳の活用を試みた翻訳者が少なくない。だが、翻訳ソフトは文脈に沿う訳を選べるような機能を備えていないこともあり、その感性を満足させる訳文にはならない。最高のソフトに巡り会えたわけでもないだろう。また対象とした文書が法文や特許文など日本語としても特殊で汎用の機械翻訳には向かなかった場合もある。さらに前編集や後編集だけでなく、その頃は受注時の形態が紙の媒体という慣習のため、スキャナーで読み取り、誤りを訂正する作業がかなり負担になる。それでも機械を有効に活用した翻訳者もいるが、トータルには人手の方が速いという判断が多くなったとしても仕方ない。

3 翻訳ソフトの品質向上

今のパソコンは購入時に翻訳ソフトが組み込まれているが、翻訳品質はソフトによって雲泥の差がある。これは93年以降開発部門を解散した企業が多く、文法が改良されていない一方で、開発体制を維持した（ソフトハウス系と電気通信系）企業数社が改良を継続したことによる。前者の企業では、サービス部門やプログラマーが語彙や付加機能を増強して後継バージョンを出すものの構文処理能力そのものが改善しているわけではない。

しかし、開発要員が文法コアの改善を重ねている翻訳ソフトはその頃から構文の解析・生成面で10～15%、訳質で13～18%ほどは向上している。分野にもよるが、文書によっては85%を超える精度になる。（同じ文書でも文章によって75～95%の変動が見られるが、これは多義性の含有率に対応する。）専門辞書を組み込めばこれまで守備範囲になかった専門分野の仕事も増やせるし、翻訳者の知識をユーザ辞書（語、連語）やパターン辞書に登録すれば、意に添う訳文が安定的に出せる。（一般企業ならそのためのカスタマイズ・サポート契約をすることもできる。）仕事に欠かせないとする翻訳者も少なくない。さらに、近年は翻訳メモリー機能を装備するソフトも増えている。ローカライゼーション業務においては翻訳メモリーの導入が急激に拡大・定着したが、これはあくまで過去の翻訳データの活用であり、新しい文書に対応できる機械翻訳システム(MT)の活用とは違う。

4 どのように翻訳が改善できるのか？

機械翻訳導入の条件としては、①翻訳品質、②編集作業の軽減を挙げることができる。

（数年遡るが）バージョン切り替えて文法が改善され品質向上の中身が伺えるケースを見よう。たとえば、下記「TransLand」（ブラザー工業）では直訳から英語ら

しい表現に変わっている。

原文：「どの先生もアメリカに行ったことがない」

Ver.4.0：No teacher has ever been to America.

←Ver.3.0：Every teacher has never been to America.

原文：「この荷物は10kgの重さがある」

Ver.4.0：This baggage weighs 10 kg.

←Ver.3.0：There is weight of 10 kg in this baggage.

次例では構造処理に相違が見られる。

原文：It is summarized in such metaphorical views of language as the dress or tool of thought.

富士通ATLAS V7：「それは[言語の][考えのドレスかツールと同じくらい]比喩的な視点でまとめられる」

富士通ATLAS V8/9：「それは[考えのドレスやツールのよう]な[言語の比喩的な視点]でまとめられます」

5 普及版と上位版で翻訳品質は違うか？

どちらも構造解析・生成能力は変わらないのだ。翻訳に違いがでるとすれば基本辞書の単語・複合語数のほか専門辞書の有無、パターン翻訳の可否に起因するものだが、一般的なテキストやホームページではほとんど差が出ない。（下例参照）。

原文：The thinking which seems to involve language is of a different kind.

ロゴヴィスタ「LogoVista XPRO 2.0」>「コリヤ英和一発翻訳バイリンガル3.0/4.0」(左が上位版)

「言語を伴うように思われる考えは異なった種類の(こと・もの)である/です。」

東芝「翻訳プロフェッショナル7.0」>「翻訳オフィス4.0」>「翻訳インターネット6.0」(左が上位版)

「言語を含むように見える思考は異なる種類であります。」

特に、専門分野の場合は、仮に構造処理に誤りがあっても、専門語・複合語に関しては訳語の信頼性が高く揺れがない。慣れない分野の英訳なら人手で探すより効率的で正確だ。このことから翻訳者の守備範囲が広がる。「MED-Transer」（クロスランゲージ）のように医薬文

書に特化したものは和訳だけではなく論文作成にも利用されているが、汎用ソフトでも専門語・複合語や用例を組み込む限り適切な翻訳が期待できる。たとえば、'aromatic compounds'は基本辞書では「香高い合成物」に訳されるが専門辞書を設定すると「芳香族化合物」になる。ただし、最近は基本辞書の拡大に伴い専門語が大量に収録され、標準設定のままでも適訳になるソフトが多い。

6

翻訳品質を支える文法・辞書情報

翻訳の精度は、基本的には、構造処理能力と語彙選択能力によって決まる。構造の解析と生成を行うのに、(1) どのような規則群すなわち文法を組み込んでいるか、(2) その規則群の動作や語彙選択に必要な意味・構造情報をどれだけ辞書に記載しているか、といったことに依存している。

たとえば、「妹は君が嫌いみたいだ」は、My sister seems to dislike you.とほぼ正しく翻訳するソフトだけでなく、You seem to dislike my sister.のように、「君が」を主語と解析する例もみられる。「好き・嫌い」など「感情」に関する述語は対象を助詞「が」で表示できるが、その言語事実を解析文法に組み込んでいたらちゃんと訳せる。

また、動詞seeはthat節を目的語とする I saw [that he was exhausted].では「分かった」、補文に不定詞句をとる I saw [him] [enter the room].では「見た」と訳すが、これは辞書に構造情報を盛り込めば訳し分けられる。

「飲む」は液体ならdrinkでいいが、スープならeat、薬ならtakeに英訳されなければならないが、目的語の意味特性に応じて適切に選択される。こうした言語処理を遂行できるシステムにするには高度な言語知識と大量の情報書き込みの労力が要求される。

7

前編集・後編集は必要か？

前編集というのは、結局MTの処理できる形に原文を書き換える作業なのだ。たとえば、「(犯人)というわけではない」などを「(犯人)ではない」に改めたり、「(説得)してみた」を「(説得)した」にすることで、「てみた」が「て見た」に誤って解析され persuaded and saw になるのを避ける。そうした作業は手間がかかる。ひらがなか漢字かで機能を識別させるなど、MTの解析能力の向上に伴い、こうした表現は正しく処理される。(文書の文字表記がこうした識別には不可欠なので、書き手は使うべき文字を適切に選択して漢字かな混じり文を作成しなければならないことになる。) 従来MTの業務使用に不可欠とされてきた前編集がかなり不要になるのだ。もちろん、正訳の比率が高くなれば後編集する部分も減る。翻訳品質の向上が編集作業の軽減に直接つながるのである。基本的には「原文をそのまま処理する」のが本筋。前編集を施すのは邪道というか本末転倒だ。前編集が必要になるのは、その文章が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からだ。

もっとも、何行にもなる文の場合は途中で切る。だが2行半までの長さなら、特に「文を切る」などの編集はせず、そのまま翻訳させる。日英翻訳の場合、(明快な文章を書く教育の欠落のためか) 原文が分かり難いことも少なくない。これを「普通に読んだら分かる表現に改める」という編集は認めて良いだろう。これはマニュアルに沿って文を書き換えるのではなく、文章を推敲するプロセスなのだ。なお、分かりやすく文章を書く訓練を受けている人が書いた英文ならば前編集は必要ない。ホームページや製品説明などはライターの手で平明に書かれており編集は要らないのだ。

ただし、翻訳の仕事のように訳文の完成度を求めるならば、後編集は必要だ。これはMT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だ。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。その際、翻訳ソフトが英和・和英辞書、国語辞書、用例集など翻訳作業に必要な各種辞書を装備し、後編集などが効率化される環境（ワークベンチ機能）が整ってきている。ぜひ活用したい。

8

機械翻訳で克服すべき問題とは

機械翻訳で克服すべき問題にはどのようなものがあるか？ 特に英語と日本語という言語差の大きい場合に問題となるのは「多義性」だが、これには「語」レベルと「構造」レベルのものがある。たとえば、修飾関係が複数あった場合に、どちらの修飾関係が正しいかを決定しなければ翻訳できないという問題がある。具体例で見てみよう。

- ① I saw a girl [with a telescope].
→ 「望遠鏡で見た」「望遠鏡を持った」
- ② 「可愛い女性の靴」
→ 「可愛い靴」「可愛い女性」
- ③ 「きれいな瞳の女性」

①の場合のどちらになるかは確率的には50%だ。②の場合、工学的には修飾語に近い名詞を被修飾語として選ぶ「近接原則」を設定できるが、統計的には65%ほどの正解にしかならない。一方、③の場合は言語的な条件で修飾関係が決定できる。「身体の一部」は誰もが持つものであり、その弁別的な特徴の記述が不可欠。「瞳の女性」では意味がないため「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。

言語的なものを含めて考えると、本来、多義性を決定するには文脈情報が必要だが、その情報がなくても、ある程度は正しい結果を得ることができる。もちろん、文脈処理機構を活用できれば問題ないが、実際それを作る

のは極めて困難で、見通しが立たない。人間の常識なり知識なりを汎用的にシステム化しようとしても、我々はそれだけの知識を持っていない。知識処理システムとかAIを、本格的に機械翻訳に組み込むようなシステムの構成は、20年後、30年後も考えにくいということだ。ただし分野によっては、ある種の文の修飾関係が、その分野だけ特異な関係で掛かるというケースもある。具体例を見よう。

【化学論文の例】

Water crystallizes to form snow.
結晶化の結果、雪が形成される

化学論文では、不定詞表現は80%以上の確率で「結果」の表現となる。そういうものを組み合わせると、本格的な知識処理システム、文脈情報処理機構がなくても、ある程度まで多義性の問題を克服出来るだろう。これは英語と日本語のように大きな言語差がある場合は問題になる部分だが、英語⇄欧州言語、日本語⇄韓国語のような極めて近い言葉の場合には、曖昧なままで翻訳が成立するので問題にならない。

曖昧性の解釈は読者やユーザがやればいいことで、(①のように) 英語の原文が曖昧でも、(J'ai vu une fille [avec un telescope].のように) フランス語にそのまま訳して構わない。原文と同じ曖昧さを持つが、解釈は読む人間がする。それで翻訳が成立する。ところが、日本語に訳すには、修飾関係を決定しないと訳そのものが出せない。それが言語差の大きい場合と少ない場合の機械翻訳における大きな違いなのだ。(語彙や構造面でも共通のところが多いことから、日韓翻訳ソフトは95%±3%ほどの翻訳率になる。)

9

英日翻訳と日英翻訳の違い

次に英→日翻訳と、日→英翻訳の言語処理上の違いについて考えよう。英日翻訳のほうが日英翻訳より10%

前後は翻訳率が高くなるが、これには理由がある。英語が「構造維持的な言語」なのに対し、日本語は「非構造維持的な言語」という反対の特徴を持つからだ。

英語は文の主要成分を代用形を立てて残すが、日本語では既知情報は残さない。このため、英日翻訳では代名詞を「それ」や「彼女」のように、そのまま訳せばいいが、日英翻訳では主語なり、目的語なりの要素を具体的に表わさなければいけない。日本語の側で主語、目的語などの要素が欠けていたら、何が欠けたかを復元しなければいけないのだ。人間が訳す場合は周りの文脈を見て、欠けた情報を復元できるが、機械翻訳では文脈情報を翻訳の際に使えない。

情報が欠けていれば翻訳できないということだが、言語処理の面でうまく対応できないわけではない。具体例を挙げよう。

「[最初に議論すべき]問題」

①a problem [to discuss first]

②the problem [which should be discussed first]

③the problem [that you should argue first]

主語として「我々」とか「あなた方」が想定されるが、原文には現れない。こういうものに対しては、言語的な仕掛け（①不定詞句、②受動文、③総称的代名詞）を上手に使って翻訳する対応が見られる。

このような言語的な対応によって、本来しなければならない「削除された成分の復元作業」を回避するのである。主要な文構成要素を欠く日本語は、英語に比べて非論理的とされることもあるが、見方を変えると、「分かっている情報を省く方が、構造維持のためにいちいち表わさなければいけない英語と比べ、より知的な言語であり、読み手、聞き手に高い知識処理を要求する言語である」という評価もできる。

ただし、機械翻訳で処理する文書の場合、文成分の欠落はそのまま翻訳精度の低下につながる事が多く、機械処理の予定のある文書については、意識的に文成分を

省略しないようにするか、文成分の欠落を指摘できる構造解析機能を持つ編集ソフトを利用して欠落成分を補充する作業を済ませた文書に改めることで、翻訳精度を格段に高めることが求められる。

2007年版の翻訳ソフト (*LogoVista X PRO; The 翻訳プロフェッショナルpremiere; ATLAS V13; PC-Transer Studio*) がどの程度の訳質になるかを以下(いずれもニュースが原文)に示す。①は主文が受動文で*when*の率いる従文が*after*節を内包する。②も受動文で補部が分詞句の知覚動詞構文だが関係節が続く。③は*のに伴い*に続く節を内包し専門用語を含む。④は*前に*節を含む連体修飾節を内包する。①、④では、「節に埋め込まれた節が(主文を修飾する)誤解析になっていないか」もチェックすべきだが、文脈処理機能を欠くことも認識して評価したい。

①The children were severely malnourished and near death last month [*when* doctors saw them [*after* social workers took them to a hospital]]. [LogoVista]: 医者が、ソーシャルワーカーが(彼・それ)らを病院に連れて行った後、(彼・それ)らを見たとき、子供たちは先月ひどく栄養不良で、そして死の間際でした。/ [The翻訳]: 子どもたちはひどく栄養不良で、ソーシャル・ワーカーが病院にそれらを持っていった後、この前の月、医者がそれらを見た時、死に近づきます。/ [PC-Transer]: ソーシャル・ワーカーが彼らを病院へ連れて行ったあと医者が彼らに会ったとき、子供たちは先月ひどく栄養失調で近い死であった。

②Black smoke was seen *pouring* from the nearby Kashiwazaki nuclear plant, [*which* was automatically shut down in the quake]. [LogoVista]: 黒い煙が震動で自動的にシャットダウンされた近くの柏崎原子力発電所から流れ出ているのを見られました。/ [The翻訳]: 黒い煙は、近くの柏崎原子力発電所から流出するのを見られました。それは震動で自動的にシャット・ダウンされました。/ [PC-Transer]: 黒い煙は近くの柏崎原子力設備から流れ出ているのを見られた。そして、

それは地震で自動的にシャットダウンされた。

③女性では、[閉経で女性ホルモンが急激に低下する]のに伴い、のぼせ、不眠といった更年期症状が現れる。

[The翻訳]:In a woman, in connection with female hormone falling rapidly, he feels dizzy with a menopause, and the climacteric symptom of insomnia appears./ [ATLAS]:For the woman, it accompanies so that the female hormone may decrease rapidly because of climacteric, it dazes, and symptoms of menopause of insomnia appear.

④ [[地震が発生する] 前に場所や時間、規模を言い当てる] 地震予知は今の技術ではできない。[The翻訳]:Before an earthquake occurs, the "earthquake prediction" which guesses a place, time, and a scale right is not made with the present technology./[ATLAS]:"Earthquake prediction" to guess the scale right cannot be done in the current technology at the place and time before the earthquake occurs.

10 翻訳の壁と克服技術

日本において電気通信系大手各社が機械翻訳システムの第1世代を構築したのが80年末だが、90年代になると、大学などにおける機械翻訳の研究は「用例ベースの翻訳」と「コーパスベースの翻訳」が主流になる。

自然言語においては、「言語形式が限られていてもその中に複数の意味を込められる」という特徴がある。これは有限の資源の有効活用という点で人間にはメリットだ。語の意味や修飾関係の多義性は文脈情報に照らせば一義に決定できる。だが、機械にはそれができない。汎用の文脈処理システムは設計デザインさえ不透明だからだ。

そこで多義解消のために出現頻度や共起情報そして用例を利用する方策が選ばれた。確かに、分野、文書の種

類を限定すれば、意味・修飾関係が一通りになるケースが多く、多義性の問題に一定の効果はある。だが、一般的な文書になると、そうした効果は期待できない。

日本における機械翻訳の研究・開発は初期段階では工学系研究者が主導してきたが、どのような文法規則が設定可能かと掘り下げることもないまま、「文法規則だけでは翻訳に限界がある」と拙速に断を下し、これに換わるブレークスルーとして用例ベースの翻訳手法に走った。「制限文法」の導入と同じ轍を踏んだのである。

—制限文法—

機械翻訳の構文処理能力がまだ幼稚な時には、この低い処理能力に合わせるべく「規格化日本語」「制限文法」といった煩雑な書き換え規則が提案された。私は「文法記述を精緻にすれば構造処理能力が上がるので制限は必要なくなる。規格化された日本語への書き換え作業は省力化を目指す機械翻訳の理念に反する。」とその誤謬を批判した。実際、主要な構造はほぼ処理できるようになり、文法制限を唱える声も消えた。

用例ベースの研究と並んで工学的手法として盛んになったのが、コーパスから言語知識を自動抽出する研究であった。

言語データ（コーパス）を形態素解析し品詞などを付与して新しい複合語を抽出したり、共起関係から統語規則を抽出する。さらに対訳テキスト（パラレルコーパス）から句単位で用例を抽出する。これを自動的に実行する技術が利用できるが、最終判断は人間が行う。この作業を適切に遂行するには、対訳を構成する表現の対応が取りやすいように整列（アライメント）しなければならないが、それはそう簡単ではない。

アカデミックな脚光は浴びないが、市販用に翻訳ソフトの性能向上に向けて文法規則の精緻化と辞書の整備を図ってきた企業では、翻訳者でも有効活用するだけ翻訳レベルの高いものを言語系スタッフが作り上げてきた。

それと較べると統計や用例ベースの翻訳は汎用性の低い実験的な翻訳の域を出ていない。90年代後半になって「文の骨格となる構文構造の処理は、用例ベースの翻訳ではうまく行かない」という認識ができたようで、(名詞句など) 文の枝葉部分に適用を限定するように軌道修正されている。

11 人間的システムの構成

そもそも、用例ベースの翻訳の基礎となる言語データは有限個の文法が産出したものである。膨大な言語データをコーパス化したり統計処理することにより用例パターンを構成したとしても、天文学的な数(60億パターンという計算も!)になる。今後も巨大化するコンピュータ・パワーで強引に処理すれば翻訳品質が改善するかもしれないが、それは人間の脳中における言語処理とは全く異質の処理だ。我々は「対訳データを記憶して、話すとき聴くときに、同じ文を検索して、訳文を引き出す」ようなことはしていない。

機械翻訳は、そのシステムの構成を設計するにあたって、人間の言語処理をコンピュータ上で実現するようなものを目標にして良いのではないだろうか。人間の言語は脳の言語中枢と関連領域(聴覚野、視覚野、前頭連合野など)との共同作業によって処理されている。文字処理、音声処理が末端にあり、中枢に辞書情報検索に支えられた構造と意味の解析・生成がある。言語的意味を文脈的に適切に理解するには知識処理領域へのアクセスもいる。現代の言語学はいろいろなアプローチの研究領域に分裂、発展してきたが、一貫して言語を「中核部に文法と辞書を装備する体系」として捉えてきた。

出現頻度や共起情報から抽出した言語規則や用例パターンは、こうした中核を補完する目的で大いに活用すればよい。標準設定の訳で不満ならば、翻訳者の知識をユーザ辞書(語、連語)やパターン辞書に登録すれば良い。我が意を得た訳になる。

12 翻訳品質の底上げ

パターン翻訳は、用例ベース翻訳の一種であるが、対訳データの中から「典型的な表現を選択してパターン登録」する点で、機械的に構築した膨大な対訳データをそのまま使用する用例ベース翻訳よりメモリーと処理速度で勝っている。「単語列の検索と置き換え」という処理を行うのだが、これは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 —

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。本来前者の記述形式をとるが、後者の記述になると限りなく文法ルール構造記述に近づく。用例パターンには、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」に訳す形式などがみられる。(なお、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」を使うと点で区別される。)

13 文処理範囲の拡大と文脈処理

現在の機械翻訳は一文単位の翻訳ということになっているが、これは「文頭から文末のピリオドまでの範囲で処理を完結する」設定にしているということだ。この文の中に幾つも文が埋め込まれることもある。そうした複合的な構造を内包する文の解析に際しては、上位構造を解析している途中でその作業を保留し、内部構造の解析

を遂行した段階で、中断した作業に戻る手続きが働く。多重並列的な処理である。

ここでピリオドという文境界の設定を外してみよう。そうすると機械は所定の範囲内の複数の文を同時に睨みながら整合性の取れる訳を出す作業をすることになるが、これは現在の技術の延長線で行える。見方を変えると「若干の先行文の情報を参照して翻訳する」ということで、「軽装の文脈処理」に相当する。

そもそも句読点やピリオドというのは、書き手が恣意的に使うもので、そこに含まれる単位文の次元でみると、本質的な違いはない。埋め込み関係を把握しながら解析する仕組みがしっかりしていれば、形式上長文1文だろうが短文3、4文だろうが、3行（120字）までの範囲なら十分対応できる。それだけの範囲を処理対象とすれば、照応関係や省略語句の推定、（先行文中の語との共起関係に基づいた）語義の選択などの文脈処理がどうか可能になるだろう。人間が行っているのはまさにこうした多重並列処理なのである。

14 悪文追放（ライティングの確立）

最近の特許文対応を謳う翻訳ソフトが増えている。英日版は英文が明示的な構造になっているのでそれなりに実用性はある。だが日本の特許文は「請求項を1文で記す」という記述条件に従い、本来数文ないし十数文から構成される文章を無理やり1文にしている。その中に含まれる節や名詞並列構造の相互関係が極めて分かりにくく、技術者でも何度か読み返さないと内容が理解できない。これを英訳するには「解読分解作業」が避けられないのだ。長い文は接続表現を手がかりに自動分割できるが、内部表現を自動的に前処理するには限界がある。特化した文法や前処理プログラムを組み込んでも、部分的な対応にしかならず、まともには機械翻訳の対象にならないのだ。敢えて機械翻訳を使うというならば、特許文はむしろ普通の文にリライトし、それから翻訳する方が

現実的だ。とにかく一刻も早く旧態依然たる記述条件を改めるべきだろう。（「特許文の改善に向けて」（『eとらんす』）パベルプレス2003.8、「特許文の現代化と機械翻訳」（Japio創立二十周年記念誌）2005、「特許文の多言語機械翻訳」（Japio 2006 YEAR BOOK）2006参照。）

15 多段階統合翻訳

今後実務に使える翻訳ソフトはどのような展開をするのだろうか。文法規則の更なる精緻化と辞書情報の充実が柱になるのは間違いない。膨大な新語や複合語などに迅速に対処するために、統計的な技術で言語情報を自動抽出するとか、用例パターンを利用してこなれた翻訳を保障する。そうした形で、工学的な技術を活用するという方向も揺るがないだろう。あくまで局所的ではあるが文脈処理機能を持たせる研究も実用化が近づいている。また、機械翻訳とは原理的に違うのだが、マニュアルの現地化などの業務にはメモリー翻訳機能も欠かせない。辞書などを備え校正などの作業が快適にできるワークベンチ的環境を提供すべく、翻訳ソフトは更に重装備化するだろう。

音声翻訳の実用化も着々と進んでいる。翻訳ソフトの解析生成文法はほとんどが書き言葉対応だが、（通訳システムに組み込む）話し言葉の文法が全く違うというものではない。コア文法は共通である。特に違うのは文末表現だが、これを部品化して組み込めば「通訳システム」になる。方言だって通訳できる。言い淀みや言い換え、「チャット」を翻訳したり、音声認識・合成機能を持たせて旅行関係の通訳をするソフトも製品化されているが、将来的には「通訳携帯」の登場も見込まれている。