（財）津田塾会設立４０周年記念日本語国際シンポジウム予稿

機械翻訳の誤りと人間の誤り

田中穂積（東京工業大学工学部）

1．言語理解と翻訳

言語学，心理学，哲学の分野で，言語とは何かを巡って様々な研究が行われている．言語学者は主として言語の形について，言い替えると統語論の観点から，哲学者は古くから言語の意味について，心理学者は言語を学習したり理解する過程に興味をもって研究してきたといえるだろう．以上の研究者の他に，人工知能の研究者が言語とは何かを巡る研究に参入してきており，その存在を無視することができなくなっている．

人工知能の研究者の興味は，コンピュータに言語を理解させることである．これが真の意味で可能かどうかはさておき，コンピュータに言語を理解させることは言語理解システムの研究と呼ばれている．これは現在でも人工知能の分野の主要な研究課題の一つである．この研究は，言語を理解する過程をコンピュータ内部にシミュレートしなくてはならず，言語に関する従来の理論のどれが有用で，何を（重点的に）研究すべきかを，言語学者，心理学者，哲学者とは異なる視点から明かにした．言語理解には膨大な知識が前提になる．彼らは，言語理解のためのこの膨大な知識をどの様な形式で計算機上に表現し，それらをどのタイミングで取り出し利用するか，理解した結果を新しい知識としてどう蓄積して再利用するかを問題にした．そして，これらの問題の解決こそが言語理解システム作成の鍵となることを明らかにした［Schank　７７］．

ところが言語理解のための膨大な知識をコンピュータに組み込むためには途方もない労力を要す［Lenat　８６］．一方我々人間は言語理解のための知識を社会生活での様々な体験を通して無意識の内に学び記憶している．意識的に知識を記憶にたたき込んでいるわけではない．コンピュータにこうした学習機構をもたせることができれば，膨大な知識を逐一コンピュータに組み込む必要がなくなる．言語理解システムの研究に関連し最近学習の問題の重要性がクローズアップされてきているのはそのためである．いずれにしても言語理解システムの研究は本質的に困難な問題を内包しており，長期の研究課題である．

ここで本題に戻り翻訳の問題について考えて見よう．翻訳はある言語を別の言語に言い替え出力することであるから，言語理解の他に文生成の問題を含む知的で総合的なプロセスであると見なせる．機械で翻訳することの難しさは，この複雑な総合プロセスをシミュレートしなければならないことである．翻訳には，言語理解と文生成に関する技術的な問題の他に翻訳可能性という原理的な問題もある．これは，言語が文化の反映であるなら，真の翻訳には二つの言語を取り巻く文化的な背景をも考える必要があるとする考え方である．もし我々が母国語でない言語を話す国の文化を知ることができないとすれば，翻訳は原理的に不可能であるということになる．しかし，たとえそうであるとしても，翻訳が意味のないことであるという結論が導き出されるわけではないことに注意したい．原理的に不可能と思われることに挑戦することは意味のあることである．

2．　人間と機械翻訳システムとの調和

さて，機械翻訳が抱える大きな問題は文化以前の問題であることは明らかだろう．現在の問題は我々自身の行う翻訳プロセスについての知見が十分に得られていない点にある．その知見が得られてたとしても，その知見をどの様にコンピュータ上にシミュレートするかということも十分には分かっていない．現在の機械翻訳システムは，我々の翻訳プロセスについての不十分な知見と，現在の自然言語処理技術を背景にして作られている．したがって，現在の機械翻訳システムはこのような限界を意識して設計されている．幾つかの商用システムが誕生しているとはいえ，現在の機械翻訳システムは未解決の多くの問題を抱えているのである．

このことは，機械翻訳結果の誤りには現在の自然言語処理技術の抱える問題が投影されているということを意味している．従って現在の機械翻訳の誤りを分析することにより，将来の機械翻訳システムが解決すべき問題が明らかになる［長尾　86］［辻井　87］［新田　87］．また自然言語処理技術のレベルアップにもつながる．

ここで次のことを注意しておきたいと思う．現在の機械翻訳結果に誤りが含まれることが自明であるとすれば，その修復は人間に委ねられることになる．人間と機械翻訳システムとの協調が必要になるのである．この協調のあり方次第で，現在の機械翻訳システムが実用になるかならないかが決まる．よりよい協調関係のあり方を設

定することは，これからの機械翻訳システムの将来に大きな影響を与えると考えられる．

よりよい協調関係をどう設定するかは機械翻訳システムの豊富な使用経験に基づき決められなければならない．よりよい協調関係の設定に向けて，忍耐強く比較的長期的な視点を持って研究しなければならない．わが国の機械翻訳システムのレベルは世界のトップクラスにあるといってよい．機械翻訳システムが商業ベースに乗り始めているのも，欧米諸国には見られない現象である［野村　８８］．さらに機械翻訳システムの使用経験に関しわが国は世界のトップにある．この優位な立場を生かし，わが国で，人間とのよりよい協調関係を持つ機械翻訳システムが誕生することを期待したい．

使用経験に基づき人間とシステムとの協調関係を練り上げる一方で，機械翻訳システムの使用者を教育する場を作ることも必要になる．これは既存のワープロ教室より高度な教育になるかも知れない．機械翻訳システムの仕組みの概略ぐらいは教育しなければならないからである．某翻訳会社で，そのような試みが既に始まっていると聞く．このことは，翻訳対象分野を特殊な技術専門分野に制限することにより，現在の機械翻訳システムが，人間との協調付きという条件で実用化に向かって着実に前進しつつあることを示している．

3．　人間の翻訳の誤り

人間する翻訳の誤りについては，翻訳家の間で様々な指摘がなされているので，以下ではやや本質的ではない特殊な誤りについて述べてみたい．

前章の最後に，特殊な技術専門分野に限定した機械翻訳は条件付きではあるが実用にもっとも近いということを述べた．このことはこの種の文献が詩歌の様に特殊な言い回しや暗示を含まないということの他に，一般の翻訳者には特殊な技術専門分野の翻訳が意外と難しいということを意味している．特殊な専門分野の翻訳に堪能な翻訳者数が限られているという事情もある．一般の翻訳者が専門分野の文書を翻訳する場合には，頻繁に現れる専門用語の翻訳に誤訳が含まれることが問題のように思われる．実際卒業研究のために筆者の研究室に入ってきたばかりの学生に，英語の技術書物，文献を読ませた乏しい経験によると，専門用語の翻訳に誤りが多々みられる．専門用語が正しく翻訳されないと（特に専門用語を知っているものから見ると）意味不明な翻訳になる．次の例を見てみよう．

翻訳例 1

(1) This is known as the constraint of referential integrity. Without this constraint, ...

(1') これはreferential integrityの束縛といわれる。この束縛がないと、・・・・。

翻訳例 2

(2) Note that the use of the character # as part of an attribute usually means that the corresponding values are identifying numbers: They may be actual numbers used in the real world (e.g. part numbers used by a manufacturer) or they may only mean somthing to .....

(2') 注意すべきことは、属性の一部としての記号#の使用は普通対応する値が識別数であることを意味することである。・・・実世界で使われている現実の数であっても良いし（例えば、生産者は部分数を用いている）、・・・。

（１），（２）共，関係データベースに関する文である．翻訳結果の（１'）では，integrity を「保全」と訳すことができなかったことは止むを得ないとして，constraint を「束縛」と訳したことが問題だろう．関係データベースの用語ではこれは「束縛」ではなく「制約」である．「束縛」という用語は理工系分野では bind の訳語として定着している．そのため翻訳結果を読む側で「束縛」に対して bind という英語を（無意識に）想定してしまうので意味不明になる．専門用語に詳しくなくその連想ができない人は，前後の文脈から constraint が「束縛」と訳されていても，文の意味をくみ取ることができるかも知れない．

次に翻訳結果の(2')をみると，関係データベースの知識があれば，最初の文中の number は「数」と訳さずに「番号」と訳すべきであると分かっただろう．部品番号は最頻出の属性名の例だからである．それでも最初の文の翻訳では number を「数」と訳したために致命的な誤りには至っていない．ところが次の文の括弧の中の part numbers の part を「部分」と訳し，numbers を先と同様「数」と訳したために「部分数」と訳してしまっている．これは明らかな誤訳である．これは「部品番号」と訳さなければならない．誤訳した当人も実はこの「部分数」の意味が分からなかったと思われる．そこで括弧の中の used 以下の翻訳も変になっている．括弧の中は「例えば、生産者が使う部品番号」とでも訳すべきだろう．

以上から言えることは人間の翻訳では，専門分野の知識不足のために専門用語がうまく訳せないこと、またその専門分野に現れる一般用語の訳語選択を誤ることがあり，その結果がその後の翻訳にも影響することがあるということである．人間は強引に辻妻合をして奇妙な翻訳をすることがある．

翻訳例 3

(3) The details of how lazy evaluation is performed efficiently and how to avoid re-evaluating functions and common sub-expressions are ... .

(3') いかに非厳密な評価が効果的に実行されるか、そして再評価する関数とコンマの副表現・・・。

文（３）では，訳者は計算機科学の分野で使う専門用語 lazy evaluation の訳語を知らず苦し紛れの訳語（「非厳密な評価」）を当てている．この専門用語はそもそも意味的には delayed evaluation とすべきだと思われるのであるが，eager evaluation の対として，日本語に翻訳するときやや誤解を生む専門用語 lazy evaluation が使われだした．したがってこの苦し紛れの翻訳では日本語として意味が正確に伝わらなくなってしまった．

人間は英文の読み間違えでとんでもない翻訳をすることがある．むろんこの種の例は多くないが（３'）では common をコンマと訳している．訳した当人はもちろん common の意味を知っている．訳者に聞いたところ，なぜこの様な翻訳をしてしまったかいま考えると見当がつかないといっていた．その時疲れていたのは事実のようだ．人間は疲れを知らない超人でもなければ，如何なる状況でも常に冷静で精神集中ができる生き物でもない．人間は疲れて精神集中力が低下しうっかり読み誤り誤訳することがある．それに対して機械は疲れを知らない．長時間の翻訳を考えたとき，機械翻訳システムが疲れを知らないということは，機械翻訳の利点の一つと考えて良い．

4. 機械翻訳の誤り

機械翻訳の誤りと言っても，個々の機械翻訳システムの性能に差があるので一概に論じるわけにはいかない．本章では機械翻訳の典型的な誤りを幾つか述べる．そしてその誤りの原因について簡単に考察する．なお本章の翻訳例は財団法人日本規格協会の調査資料を一部参考にしている［日本規格協会　８８］．

翻訳例 a

(a)　　１１枚型磁気ディスクパック

(a')　　11 type magnetic disk pack

正確な専門用語は magnetic eleven-disk pack である．おそらくこの専門用語を登録しておかなければ正確な翻訳はできないだろう．しかし枚数が１０枚，１２枚，．．．と異なる場合，ten-disk, twelve-disk,..のように異なる専門用語を登録しておくのだろうか．それもやむを得ないとする考え方もある．しかしできればそれを避ける方法を見つけたいものである．

翻訳例 b

(b)　　スピンドルロックショルダ

(b')　　spin dollar rock shoulder

形態素解析の誤りのためにこのような結果が得られた．日本語では専門用語とそうでない語との間に切れ目がないのでこの様になったとも考えられる．日本語では仮名書きのロックという用語も，英語ではrock, lockの２通りの意味があり曖昧である．そのいずれが正しいかという問題もある．ここではスピンドルロックが専門用語で，spindle lock を辞書に登録しておいたとしよう．その場合に，shoulder of the spindle lock という正しい翻訳が得られるだろうか．おそらく the spibdle lock shoulder という翻訳結果が得られるのではないだろうか．

翻訳例 c

(c)　　記録面及び記録面の番号

(c')　　number of recording surface and recording surface

意味的には「記録面」と「記録面の番号」であろう．翻訳結果もそう解釈できないわけではない．しかし翻訳結果の第一義的な読みは[number of [recording surface and recording surface]]であろう．これは意味をなさない訳である．並列表現は現在の機械翻訳がうまく扱えない典型的な例である．発見的なその場しのぎの解決策はあるにしても限界がある．並列表現の処理には意味や常識が必要になることが多いからである．しかし意味や常識を用いた自然言語処理は十分に研究がなされていない．

並列表現に対して直感的には意味的に近いものを並列させるとい

う戦略が考えられる．たとえば

東洋の仏教と西洋のキリスト教

という文では「仏教」と「西洋」という対と，「仏教」と「キリスト教」という対を比べると，後者の対は共に「宗教」で前者より意味的に近いから並立する．しかしこの考え方では文（ｃ）はうまく扱えない．なぜなら文（ｃ）に現れる２つの「記録面」は意味的に近いどころか同じだから並立してしまう．この問題を助詞「と」の前後の同一名詞は並立させないとするヒューリスティックで解決するのが最善かどうか．様々な文例に当たって検討する必要があろう．助詞「と」や「及び」などで並列した表現の解釈は、翻訳者ならほとんど誤らない．機械による翻訳と人間の翻訳の違いがよく現れる例だと思われる．

名詞句ではなく文が幾つも並んだ場合も問題を引き起こす．たとえば「Ａし、Ｂし、Ｃした後、Ｄする」という文が，Ｄ after Ａ，Ｂ，Ｃのように訳され，順序に曖昧性がでて分かりにくくなることがある．

翻訳例ｄ
(d)　一般事項　JIS K7701 の２による．
(d')　General.item　By 2 of JIS K7001

文（ｄ）には主語が省略されている．そのため翻訳結果が意味をなさない．正確には「一般事項は」という主語を補った翻訳を行うべきである．省略語の補強は機械翻訳では最も困難な仕事の一つである．専門家でないと省略語がどれであるかを判別することが困難なこともある．専門家にできないことが機械翻訳システムにできないことは当然であるが，専門家でない人が容易に判別できる省略語が機械翻訳システムにできないことが多いのは問題である．今後の重要な研究課題であろう．なお(d)は(d'')のように訳さなければならない．

(d'') General item　The general items shall be in accordance with 2 of JIS K7701

翻訳例ｅ
(e)　水200mlにより化カリウム(KI)120gを溶解して得られた溶液によう素2.7gを加え、・・・

(e')Add 12.7g is added to an obtained solution by dissolving the potassium of (KI) 120g iodation to 200ml water, ...

翻訳例(e')では、原文の[[・・・溶解して得られた]溶液]とすべきところを[・・・溶解して[得られた溶液]]のようにくくって翻訳している. 長い文で埋め込み文を持つものは一般に多数の統語解析結果が得られるが, そのうちどれが妥当かを決めるためには意味的また文脈的な情報を使わなければならない. 現在の翻訳システムにとってこれは困難な仕事であることは既に何度も指摘してきた.

次の例はおそらく現在の機械翻訳システムにとって難しいと筆者が想定したものである. 対になっている翻訳結果は想定機械翻訳結果である.

翻訳例 f

(f) John brought a tool box.
He picked up the jack.

(f') ジョンは道具箱を持ってきた.
彼はジャックを取り上げた.

翻訳例(f')は道具箱の中にあるものを取り出すわけだから, jackは「ジャッキ」と訳すべきだろう. このような訳語選択には前文からの情報(文脈情報)が必要になるだけでなく, 道具箱の中には普通どのようなものが入っているかという常識が必要になる. 道具箱の中にはカードの「ジャック」は普通入っていない. 「ジャッキ」である.

翻訳例 g

(g) The film was exhausted.
I entered my dark room.
I developed it.

(g') フィルムが尽きた.
私は暗室に入った.
私はそれを開発した.

この文では最後の指示代名詞 it の先行詞が film であると認識しない限り, develop を「開発する」と訳してしまうだろう. developの訳語選択には目的語itが何であるかを知らなければならない. 照応指示の問題を解かなければならないのである. 文内の照応

指示関係の分析法として，C＿統率理論などがあるが，これは前文にある先行詞を決めるわけだから応用できない．CarterはC＿統率理論の問題点を指摘し，文内の照応関係を解決するのに，Sidnerらの焦点モデルによる解析を優先させ，それでも複数の先行詞（後行名詞）が残る場合にC＿統率理論，そして最後に意味や常識を用いるべきだと主張している［Carter　８７］．(g)の翻訳は彼の理論を用いると可能になるかも知れない．焦点モデルについては日本語にも応用可能な理論だと思われる．

5.　機械翻訳の高度化に向けて

前章で述べたことの他に現在の機械翻訳にとって難しいと思われる問題を幾つか指摘する．第１に現在の機械翻訳システムは，翻訳しようとする文の統語解析結果の構造に強く束縛された訳文を作り出すことしかできない［新田　８７］．人間の翻訳では，１文を複数個の文にして訳出したり，逆に複数個の文を一つにして訳出することがある．それにより翻訳結果が分かりやすくまた読み易くなる．ところが（特殊な仕掛でそうなる場合を除き）機械翻訳では１文を１文にしか翻訳できない．この問題は適度な代名詞化と省略を含み人間にとって分かりやすい文体の文を作り出す問題とともに今後研究されねばならない．

つぎに英語を日本語に翻訳するとき，無生物主語の翻訳をどうするかという問題がある．無生物主語の日本語文は一般に不自然になるからである．そのほかに，一方の言語の文をそのまま訳すと不自然になるため，適当な語を補わねばならないこともある．たとえば ...enjoys his coleagues は「同僚を楽しむ（？）」ではなく「同僚との付き合いを楽しむ」と訳さなければならない．この種の問題については文献［田中　８４］を参照されたい．言語対照研究が必要になることは明かだろう．

最後にもう一つ問題を挙げるとすれば比喩を含む文の翻訳がある．我々の比喩理解の機構については未だ十分に解明されていない．したがってその機械翻訳がうまく行かないのは当然とも言える．しかし比喩を含む文の数は非常に多い［Lakoff　８３］．現在の機械翻訳のレベルをもう一段向上させるためには，今後この問題の解決に真剣に取り組まなくてはならないだろう．少なくとも提喩の理解には，部分／全体関係，上位／下位関係などの隣接関係についてのシソーラスが必要になる［田中　８７］［山梨　８８］．

参考文献

[Carter 87] Carter, D.: Interpreting Anaphora in Natural Language Text, Ellis-Horwood (1987).

[日本規格協会 88] 国際規格との整合化促進のための調査報告書, 日本規格協会 (1988)

[Lakoff 83] Lakoff, G. et. al.: Metaphors and We Live By Univ of Chicago Press(1983).

[Lenat 86] Lenat, D. et. al.: CYC: Using Common Sense Knowledge to Overcom Britlleness and Knowledge Acquisition Bottlenecks, The AI Magazine, 6, 4, 65-85 (1986).

[長尾 86] 長尾 真: 機械翻訳はどこまで可能か, 岩波書店 (1986).

[新田 87] 新田義彦 (他): 機械翻訳のむずかしさと課題, 月刊言語, 大修館, 50-57 (1987).

[野村 88] 野村浩郷 (他編): 機械翻訳, bit特集号 (1988)

[Schank 77] Schank, R.C. et. al.: Scripts, Plans, Goals and Understanding, Erlbaum (1977).

[田中 84] 田中穂積 (他): より自然な翻訳へのアプローチ [I] ――英日翻訳における表現の対応――, ICOT TR-073, ICOT (1984).

[田中 87] 田中穂積 (他): 上位／下位関係シソーラス ISAMAPの作成 [I], [II], 情報処理学会自然言語処理研究会資料, (1987).

[辻井 87] 辻井潤一: 機械翻訳のための文法とその問題点, 第一回「大学と科学」公開シンポジウム「日本語の特性と機械翻訳」, 133－156 (1987).

[山梨 88] 山梨正明: 比喩と理解, 認知科学選書17, 東京大学出版会 (1988).