

コンピュータによる異言語人とのコミュニケーション Mutual communication using our own language through computers

岩垣 守彦

IWAGAKI, Morihiko

I earnestly hope that a world may come into existence where everybody speaks his or her own language to communicate with each other: When a young man wants to befriend another young man from a different culture and in a different language, both of them can communicate in their own language as smoothly as if they were people in the same country.

As a tool to assist in the realization of this alchemical hope, I'd like to use computers, whose software is based on a series of devices common to all languages of the world.

But my hope has not yet been realized, even by MT software. In fact, we cannot translate English to Japanese effectively, much less Japanese to English.

Therefore, I'd like to make a proposal concerning the idea of converting English into Japanese and Japanese into English, in information-unit order.

ロボットを連れて散歩に出たら、知り合いのアメリカ人が歩いてきて「ハーイ」と声をかけてきた。ちょっと返答が遅れると、連れのロボットは「やー、だって」と僕の方を向いて言った……というような情景を以前から夢見ているのであるが、なかなか実現しない。

年をとると視力が落ちて本を読むのが大変な作業になる。外国語は一層やっかいである。会話は難しいとしても、外国語の資料を読んで正しく日本語に変換し、それを音声で伝えてくれるロボットなど書斎に置いておきたいと夢見ている。しかし、これもなかなか難しいようである。

機械翻訳が考えられてから60年ほど経つ。いろいろな方法が試されてきた。トランスファー方式、アナロジー方式、統計翻訳、あるいはいくつかの方式を組み合わせたものなど。しかし、まだ安心して使うことのできる機械翻訳はない。

機械翻訳が話題になったころ、MIT のパパー博士はコンピュータによる機械翻訳は不可能だ、それに費やす時間とお金があるならバイリンガルを育成に使ったほうがよいと言った[室謙二]そうであるから、現在、機械翻訳がなかなかうまく働かないのも当然と言ってもいいかもしれない。

しかし、インターネットが普及し、人々のコミュニケーションの範囲が世界的規模になってきた。そして英語が世界の共通語として使われる度合いが高まっている。それに伴って、世界から一つまた一つと言語が消滅し、その速度がますます速くなっている。現在、一年に500言語ずつ地球上から消滅しているようである。コミュニケーションの手段として、母国語より英語を使うほうが世界が広がるからである。日本では「国際人になるため」に10年も英語の学習に時間を割いている。これから小学校でも英語の勉強が始まるとのことで、更に長い時間をかけて英語を学習することになる。アメリカやイギリスの子供たちが母国語(英語)を通して様々な事に関心をもって若い頭脳を鍛えている時に、日本の子供たちは異言語(英語)に精力を注ぎ、へとへとに疲れて、挫折感にさいなまれ、言語的劣等感を一生持って生きなければならないかもしれないのである。かりに英語学習が効果をあげても、母国語をおろそかにするということは、自分の文

化的な基盤を失うということである。

世界の人々が自分の言葉で考え、語り、書き、異言語の人に伝える時には「翻訳機械」を通して難なくコミュニケーションをすることができるという状況を実現してほしいと願っている。

しかし、いまだに「機械翻訳」は十分ではない。これは、たぶん、言語に対する理解が昔とほとんど変わっていないからである。コンピュータのメモリーやハードディスク、演算速度などは、機械翻訳が試みられた当時には考えられなかったほど廉価化、多量化、高速化などが進んだ。それにもなってコーパス言語学が誕生した。言語の資源化、統計的处理には寄与しているが、この言語学も「言語理解」を変えたり深めたりしてはいない。たとえば、

I was walking along the street, when I met Tom.

と

When I was walking along the street, I happened to meet Tom. の違いを説明することはできない。この現象に関して、[齊藤俊雄(編)1998]によると、Greenbaum & Nelson (1996) (“positions of Adverbial Clauses in British English,” World Englishes, 15, pp.69-81) は、副詞節の位置関係の調査結果(文頭 29.3%, 文中 1.8%, 文末 68.8%)を挙げ、さらに「When 節は文末を好み(62.8%), 他方 if 節は文頭に起こることが多い(62.7%)。特に文頭の if 節については、条件・結果という意味的に自然な順序が他言語でも観察される普遍的傾向であるという」(p.125) と説明されているが、副詞節の位置関係は「好み」で決まるものではない。「情報の比重」で決まるのである。

パパー博士が否定した機械翻訳に挑戦するには、博士が当時認識していたのとは異なる言語理解に基づかなければなるまい。しかし、現在の日英両語の機械翻訳は、明治以来の「英文解釈法」の上に乗って進められている。たとえば、

There was once a Goldfish who lived in the sea in the days when all fishes lived there.

という「複文」を機械翻訳にかけると

「魚がすべてそこに生きていた時期で海に住んでいた金魚が以前いました」

と、昔の英語教師が訳すように「訳し上げ」ている。

また、連体節を含む日本文

「ニューヨークに四年ほど住んだことのある私には国際人になることがどんなにむずかしいかわかる」
を機械翻訳にかけると

*I who had lived in New York for about four years know well whether it is difficult to become an internationally-minded person.

という非文を作ってしまう。

それでも日ごと改良が重ねられているらしく、翻訳の精度はかなり上がってきている(以前は「国際人」を an international man と誤訳していた)。[柴田勝征(2006b)]は「近年の商用機械翻訳ソフトの応力アップには目を見張るものがある。完全に人間のプロの翻訳者のレベルを超えている。(中略)最近では、東大入試の和文英訳テストで60点を取った翻訳ソフトがあった。」と報告しているし、[成田一(1997)]も機械翻訳を好意的に見ている、しかし、上の二つの訳例で見ると、「複文」の変換が従来の方法ではこれ以上正しい訳に進まないと考えて、新しい「重文・複文の変換の方法」を試みている人([池原悟(2005)])もいる。従来の翻訳ソフトの欠陥に関しては、すでに[柴田勝征(1996)]が「文脈解析と機械翻訳」の中で述べている。

機械翻訳で使われている「英文解釈法」は、漢文を返り点を使って日本語として読むのと同じで、「英語を日本語として読む」方法をルール化したものである。人間は学習したルールの上に経験的知識を「パターン」として積み重ねて、自らルールの破綻を補いながら「文のパターン」を推定して運用する。先の池原氏のグループはこの「文の類型パターン」を集積して「重文・複文」をなんとか正しく翻訳しようと工夫している。しかし、実際にそのアイデアが実用に供されるまでには時間がかかりそうである。

言語による情報を見直して、コンピュータ用に何か別の「言語による情報伝達のルール」を再構築することはできないだろうか。そもそも言葉による情報伝達には何が重要であろうか。

ここに、一枚の「空間」があるとす。その「空間」に現実のものを入れる。

無機物--太陽、雲、湖、地面、岩、など--を入れる。

次に

有機物--植物--草、木、などを入れる。また、動物--魚、両生類、鳥類、哺乳類、など--を入れる。

ものにはそれぞれ「形」と「名前」があって「空間」の中で止まっている。

この「空間」に「時間」を加えてみよう。雲が流れはじめ、動物たちが動きだす。小鳥が「チイチイ」鳴き、子犬が「キャンキャン」と吠える。「時間」とは「動き」と「声」である。「時間」(動詞)が加わって「形」(名詞)が動き「事象」(名詞+動詞)が始まる。

そして、言語にとって情報伝達の要は「動詞」である。たとえ

“I love you.”

という英語では、love にストレスがあり、この love は動詞である。フランス語では、語順は変わるが

“Je t'aime.”

でも aime にストレスがある。aime は動詞である。日本語の場合は、最も極端で、

「好きです」

のように、動詞だけになってしまう。これは「言葉は動詞を中心にして一つの単位情報が形成される」ということである。情報が多くなると動詞が増えていく。先に挙げた

There was once a Goldfish who lived in the sea in the days when all fishes lived there.

という文には、was; lived; lived と三つの動詞がある。ということは、この文は「三つの単位情報と二つのつなぎから成り立っている」ことになる。前から順に単位情報を日本語に変換していくと

There was once a Goldfish(昔、金魚がいました)
who(=he) lived in the sea(その金魚は海に住んでいました)
all fishes lived there(お魚はみんな海に住んでいました)

となる。これらの単位情報が、who; in the days when という二つの「つなぎ」で結ばれて

昔、金魚がいました。それは海に住んでいました。その頃お魚はみんな海に住んでいたのです。

という情報を伝えているのである。

このような「動詞を中心にした情報の伝達」は世界のどの言語にも共通している。日本語も同じである。次の文は五木寛之の小説の一部である。

波留子はあらかじめ電話で連絡を受けていたので、夕食の支度のための買物をすませ、部屋を掃除して布由子の帰りを待っていた。(五木寛之『四季・波留子』)

この日本文を動詞を中心にした単位情報によって分けると、次の通りである。

波留子は前もって電話で連絡を受けていた
彼女は夕食の買い物をした
彼女は部屋を掃除した
彼女は布由子が帰宅するのを待っていた

日本文の「単位情報」を、それに相当する英語の「単位情報」に変えると、次のようになる。

Haruko was warned in advance by telephone
she did the shopping for dinner
she cleaned the room
she was in readiness for Fuyuko's homecoming

これで情報は正しく伝わる。優れたプロの翻訳家は、これに、日本文の「つなぎ」(ので; …, …して)に相当する英語の「つなぎ」を加え、基準になる時に合わせて述語動詞の型を決めて、もとの日本文と情報(イメージ)の順序が同じになるように

Haruko, who had been warned in advance by telephone, had done the shopping for dinner and cleaned the room in readiness for Fuyuko's homecoming.

と四つの情報をつなぐ。これは元の日本文とほぼ等価の変換文になっている。

ということは、日本語のように英語から離れていると思われる言語([Pinker(1994)]は同族ととらえているが)でも「動詞を中心にして作る単位情報」を、単位情報の順に英語化していくと情報を正しく移すことが可能であるということである。どの言語においても複数の事象は

事象(名詞+動詞型)+「つなぎ」+事象(名詞+動詞型)

という形で情報の伝達がなされ、これはだいたい3歳児の言語生活であるが、情報(意味・イメージ)が正しく伝わるなら、それでもいいのではないだろうか。

ロボット研究は日本が最も進んでいるそうであるが、それでも目標は3歳児の人間の能力だそうである。[立花隆]翻訳ロボットも3歳児とすれば、情報が正しく伝わるなら少々ごちない英語やごちない日本語も許してやれると思う。

日本語と英語では、言語の仕組みが異なるので、機械翻訳で正しく変換することは容易ではない。一番難しいのは日本文の文中に現れる連体節の処理である。英語に変換する際には関係代名詞節に訳さざるを得ないことが多く、どうしても情報の順序が逆になる。たとえば、

1949年に二度目の渡仏をした藤田嗣治は、その年のうちに、レストランにひとりで座っている女性を描いた。

を機械翻訳にかけると

Fujita to whom Tsuguharu Fujita did the second 渡仏 in 1949 drew the woman who is sitting on the restaurant alone on the inside of the year.)

という意味の通じない英文を提示する。最初の連体節「1949年に二度目の渡仏をした藤田嗣治は」は、プロの翻訳家は

Visiting France for the second time in 1949, Fujita Tsuguji, ... のように、日本語の情報順に変換するのであるが、次に出てくる「レストランにひとりで座っている女性を描いた」は

painted a picture of a woman sitting alone in a restaurant のように情報の順序を変えて変換せざるをえない。しかし、動詞中心の情報単位にすれば

Fujita Tuguharu visited France for the second time in 1949.

He painted a picture of a woman within the same year.

She sst alone in a restaurant.

と情報を伝えることができる、多くの日本文も英文も、動詞の(意味上の)主語に留意して変換すれば、前から順に情報を得ることができるのである。

動詞を中心にした情報を出来るだけ前から順に変換するという方法は、今まで先人が考え出した英文解釈の資源を無駄にすることなく、日英両文の情報を等価的に変換することが可能になるばかりでなく、単位情報順に発話された、あるいは、書かれたものを、単位情報順に理解するという意味で、世界の多くの言語を一つの共通なルールで処理することも可能になるのではないかと期待することができる。

「動詞を中心にした単位情報」で情報の相互伝達をするという観点からみると、英語は実に簡単で有機的なルールからできている。時間とその表し方を例にとると、情報は、基本的には、次の四つである：

- 「いつものこと」
- 「今現在のこと」
- 「すんでしまったこと」
- 「今以後のこと」

英語では、これらを上から順に

- 現在時制
- 現在進行形
- 過去時制
- 未来形

という述語の形で使う。さらに準動詞が同じように「情報」に対応している。

- 基準になる時 と同時 現在分詞
- より前 動名詞, 過去分詞
- より後 to 不定詞

という感覚で使われる。たとえば、

He went to London to study English.

は He went to London という「時」を基準にして、それより後に to study English をするという情報を提供している。また

Look at a baby sleeping in the cradle.

においては look at の同時点で baby が sleeping しているという情報を提示しているのである。

英語には、上で見ると「時制」は二つしかなく、これは純粹に客観的な事象の伝達に使われる。「形」で示されている「現在進行形」と「未来形」は本来は主観的な事象の伝達に使われるものであるが、それが借用されている。「主観的」と言えば、助動詞もすべて主観的なもので、上の四つの情報パターンに対応しているのであるが、それはそれぞれ客観的な事象を伝達するための伝達形を持っている。

- may be allowed to
- can be able to
- must have to
- should ought to
- would used to

など。

また「動詞を中心にした単位情報」で情報の相互伝達をするという観点から日本文と英文を見ると、どちらも、

「事象 + つなぎ + 事象 + つなぎ + 事象・・・」

になっている。しかし、そのつなぎ方は、たとえば、

「昨日、街に行きました。通りを歩いている時に、花子に出会いました」

という日本文の第2文は

{事象(私は通りを歩いていた) + つなぎ(時に)} + 事象(私は花子に出会った)

という構造になっている。日本文では、「つなぎ」が文の終わりにきて、次の文が続くという形しかないが、この日本文に相当する英語は

A. I went downtown yesterday. I was walking along the street(,) when I met Hanako.

B. I went downtown yesterday. When I was walking along the street, I happened to meet Hanako.

となる。つまり、

A. 事象(I was walking along the street) + {つなぎ(when) + 事象(I met Hanako)}

B. {つなぎ(when) + 事象(I was walking along the street)} + 事象(I happened to meet Hanako).

という構造が可能である。

しかし、AとBの違いは情報の比重の違いである。Bのように When を前につけると、英語ではその節の伝達比重が前後の文より軽くなるのである。したがって、どの事象も同じ比重で伝達すると決めれば、条件節以外は、日英両語は

日本文 {事象 + つなぎ} + 事象

英文 事象 + {つなぎ + 事象}

と対応させるだけですむ。(なお、ここで言う「つなぎ」とは、「動詞を中心とした一つの事象をつなぐもの」のことで、接続詞、関係詞のほか、準動詞も含まれる。)

経験を積んで、英文を自由に読んだり書いたり出来るようになると、私たちは昔中学校や高等学校で習った「英文解釈法」を使わない。私たちは無意識に情報を動詞を中心にした単位

情報に還元してとらえ、前から順に理解している。たとえば、次の新聞の記事、

It was almost midnight on a Saturday evening in December 1994 when General Colin Powell received a telephone call from President Clinton asking the former chairman of the Joint Chiefs of Staff to drop into the White House for a chat.
を、コンピュータのように

一般 Colin Powell がクリントン大統領から電話呼び出しを受け取った 1994 年 12 月に、土曜日の夕方、雑談のためのホワイトハウスに立ち寄るために統合参謀本部に前者の議長を要求することが、ほとんど真夜中でした。

と読む人はいない。多くの人は、たぶん、

It was almost midnight on a Saturday evening in December 1994
General Colin Powell received a telephone call from President Clinton
Clinton asked the former chairman of the Joint Chiefs of Staff
Clinton wanted him) to drop into the White House for a chat.
と動詞を中心にした四つの単位情報を頭で処理して、

1994 年の 12 月ある土曜日の真夜中ごろのこと、
コリン・パウエル将軍はクリントン大統領からの電話を受けた。
クリントン大統領は前統合参謀本部議長に求めたのは、
ホワイトハウスにちょっと立ちよっておしゃべりしていかないか
ということだった。

と理解しているはずである。実際、上の文を単文化して機械委
訳にかけると

それは土曜の 1994 年 12 月の夜にほとんど真夜中でした。コ
リン・パウエル将軍はクリントン大統領から通話を受け取りました。
彼は、雑談のためにホワイトハウスに立ち寄ってくれるように統
合参謀本部の元議長に依頼しました。

のように、現在の機械訳でも正確な情報を得ることができる
のである。

言語と言うものは、英語も日本語も、前から順に書かれるもの
であり、前から順に「単位情報化」(意味の塊・音の塊)して読む
ものである。私たちが普段頭で行っている「文を前から単位情報
順に読む」ルールをコンピュータに教えてやることはできないの
だろうか。

[池原悟(2005)]によると「日英機械訳システム ALT-J/E(池
原ほか 1987)において、単文構造を対象とした表現意味辞典
「日本語語彙体系」(池原ほか 1997)が開発されている。この辞
書は、動詞と格要素との関係を 1,7000 件の結合価にまとめたも
のである。格要素となる名詞の意味的な用法は、名詞意味属性
(2700 分類)を用いて指定されており、40 万語の名詞意味辞書
と連動する。訳訳実験では、IPAL (IPA 1987) の例文 5000 件
に対して、90%の正解率が得られており、方式限界は 97%であ
ることが報告されている。(金出地ほか 2001)これにより、日英機
械訳における単文の意味的な訳し分けの問題はほぼ解決でき
たとみられるので... (pp.1-2)とのことである。この単文訳に
「文脈解析」に優れている[柴田勝征(2006a)]の提唱するアイ
ディアを組み合わせると、かなり正確な情報伝達が機械を通し
て可能になるのではないと思われる。

もっとも、上記の二つの研究は旧来の「英文解釈法」の言語
理解をベースにしている。「情報単位」で言語理解する新し
いルールで手直しをする必要があるが、英語を「情報」という観
点からルール化して提供しなかったのは言語学者の怠慢である
し、コンピュータにプロの訳家と同じような変換文を求めたの
は工学技術者の楽天性である。早く両者が協力しあわないと、
世界からいくつも言語が消えてしまう。ということは文化が消え
ていくことである。

参考文献

- [Ferguson(1964)], Charles A.: Baby Talk in Six Languages
(American Anthropologist: vol.66 special issue: "The
Ethnography of Communication" pp.103-114)
[Taylor(1991)], John R.: Linguistic Categorization, 1991,
Oxford Univ. Press.
[Pinker(1994)], Steven: The Language Instinct, 1994, William
Morrow and Company, Inc.
[Pinker(1997)], Steven: How the Mind
Works, 1997, W.W. Norton & Company.
[池原悟(2005)]: 「非線形性に着目した言語表現モデルと重文
と複文に対するパターン辞書の開発」(第 10 回 LACE 研究
会(2005.12.24-25) pp.1-2
[岩垣守彦(1993-a)] 『英語の言語感覚----ルイちゃんの英文
法』(玉川大学出版部, 1993)
[岩垣守彦(1993-b)] 『日本人に共通する和文英訳のミス』(ジャ
パンタイムズ, 1993)
[岩垣守彦(1994)] 『よい英文を書くための和文英訳のテクニッ
ク』(ジャパンタイムズ, 1994)
[岩垣守彦(1996)] 『辞書ではわからない英語の使い方』(ジャ
パンタイムズ, 1996)
[岩垣守彦(2003a)] 『機械訳の精度を上げるための構文解
析の提案』(2003/03/25 自然言語処理研究会, 東京工科大
学)
[岩垣守彦(2003b)] 『日本文の「つなぎ」と英文の「つなぎ」の対
応に関して』(2003/05/26 自然言語処理研究会, 東京工業
大学)
[岩垣守彦(2003c)]: An Idea of Translation of Japanese contact-
clauses into English (2003/08/26 第一回アジア辞書学会, 明
海大学)
[岩垣守彦(2003d)] 『連体修飾節の英訳に関して』(2003/07/25
自然言語処理研究会, 山形大学)
[萩原裕子(1998)] 『脳にいだむ言語学』(岩波科学ライブラリー
59, 1998)
[斎藤俊雄(編)1998] 『英語コーパス言語学』(1998, 研究社)
[佐藤理史(1997)] 『アナロジーによる機械訳』(日本認知科
学会編, 共立出版, 1997)
[柴田勝征(1996)] 『文脈解析と機械訳』(福岡言語学会例会
報告, 1996/12/15)
[柴田勝征(2006a)] 『US 式和英訳システムの制作』(ラッセル
社, 2006)
[柴田勝征(2006b)] 『CREST 文型パターンと US 式文型パター
ンの比較検討』(CREST 平成 17 年度末報告会, 2006/3/21,
教育会館)
[成田一(1997)] 『パソコン訳の世界』(講談社現代新書,
1997)
[リチャード・A・ウイルソン] 著・渡部昇一・土屋典生訳(1981)
『言語という名の奇跡』(大修館書店, 1981)