インターネット翻訳による情報受発信

横山晶一

▶1　はじめに

インターネットの起源は，1969年に開始されたARPAネットや，1980年に始まったCSネットなどにさかのぼることができると言われている（村井，1995）。アメリカでこの用語が使われ始めたのは1990年代初め，日本では1990年代半ばである（岩谷，1995）。インターネットという名前のついた本が売り出されたのもまさにこの時期である。

しかしながら，一般にインターネットという言葉が喧伝され始めたのは，ここ１～２年のことである。合わせて，インターネットが実際に爆発的に普及している。日本におけるインターネット人口については，諸説あるが，ある調査によれば，1999年の初めに1500万人あまりであったのが，2000年３月には，2100万人に達している（(AAMT, 2000)の執筆途中のデータから引用。元の出典はCommerceNet)。また，携帯電話等のモバイル端末からの利用（たとえばNTTドコモのiモードなどの加入者）が，2000年夏に1000万人を突破したことも影響して，2000年末現在のインターネット利用人口は予想よりも２年ほど早く，4000万人近くになったといわれている。少し前の予測では，2005年には，人口の６割に当たる7600万人がインターネットを利用する（その半分弱は携帯電話から）とされていたが，これも２～３年予測よりも早まるかもしれない。

アジア各国を見ると，ここ２年ほどの間に，ネットワークのインフラストラクチャが整備されつつあるようで，発展のスピードは日本よりも急速である。上記のデータでは，中国では，1998年には60万人余りで，人口のわずか0.005%しか使用人口がなかったが，2000年１月には900万人弱（人口比0.7%）と，わずか２年で10数倍の利用者増となっている。また韓国も，1997年に70万人（人口比1.5%）だった使用人口が，1999年10月には950万人弱（人口比20.3%）と，これも同様の増加ぶりである。

このように，インターネットが，すでに電気や水道などと同じインフラストラクチャとして定着しつつある現状においては，日常生活でインターネットを使うということがごく普通に行なわれるのは言うまでもない。我々は，日頃生活している社会と，インターネット上のある程度バーチャルな社会とで，二重生活を送っているといっても過言ではない。つまり，家庭にいるときも，出勤して職場にいるときも，あるいは出勤途上でさえ，現実の社会とインターネット社会との並行，二重的な社会に生きている訳である。

ネットワーク社会では，持てる者と持たざる者との情報格差や一部への情報偏在，また情報過多による必要情報の隠蔽や見落としなど，負の部分も存在する。また，いきなり世界とつながってしまうわけであるから，犯罪に巻き込まれたり，知らないうちに自分が人に迷惑をかけるといったことも起こりうる。しかしながら，一般の人が全世界と即時に，しかも比較的地域差が少ない状態で接続できるという時代は，おおむね歓迎すべきものと考えられる。

インターネットで世界とつながったときに最も問題になるのは，自分と世界とのコミュニケーションをどのようにとるかということである。日本国内でメールのやり取りをしたり，ホームページ（Web上での表紙のことであるが，すでに定着しているので本稿ではこの表現を用いる）を見たり，情報検索をしたりする場合には，日本語で十分である。たまに文字化けが起こることはあっても，今日のソフトウェアでは，日本語を読み書きするのにさほど不自由はない。また，海外とも，在外の日本人や，日本語に堪能な外国人とのやり取りは，日本語を使って済ますことができる。日本語入出力としては，多少の不便さは内蔵しつつも，かな（ローマ字）漢字変換方式が定着し，かつてのように種々の入出力が混在している状況（横山，1982）はほとんど見られない。

一方，上記以外の海外とのコンタクトは，現時点では英語で行なわれることが非常に多い。英語が共通語として使用（鈴木，1999）されるのは，インターネットにおいても同様である。情報発信のためには，英語を使用することが便利であるといえる。しかしながら，インターネットでの情報受発信のために英語を使わなければいけないからという理由を，英語公用語論の論拠の一つとすること（たとえば船橋，2000）には反対である。機械翻訳のような技術の進歩によって，これが克服される可能性についても前掲書には言及されているが，現在の技術でも，機械翻訳を利用することによって，情報受発信がかなりの程度できるというのが本稿の趣旨である。

ここでは，インターネット上のホームページやメール等で，情報を受発信する場面を考え，それらの場面において，機械翻訳を利用することによって，言語の壁というものをある程度克服できる可能性について考察する。なお，ある小規模のワークショップで，本稿と同様の趣旨での発表を行ったことがある（横山，2000c）が，ここではその稿をほとんど全面的に書き換えている。ただし，内容そのものには大きな変更はない。

▶ 2　コンピュータ上での英語の役割

コンピュータは，初期のものが英語圏で作られたことや，記憶容量，それに伴う文字量の制約から，英語以外の入出力を受け付けないという，まさに英語独占の時代が長く続いた（横山，2000b参照）。日本語入出力については，1978年頃にJISの制定や，日本語ワードプロセッサの発売などを契機として，コンピュータ上で日本語を扱う動きが高まっていった。それまでの日本の実情は，コンピュータ上では，ローマンアルファベット（英語）で十分であり，英語以外の文字や言語（ドイツ語やフランス語でさえも）を扱うことによって，コンピュータ上の容量を余分に使ったり，操作を複雑にしたりする必要はないという極端な意見さえ聞かれたのであった。したがって，今日では歴史的なことであるが，独自のコードを持ったオフラインの（非常に高価な）日本語プリンタが存在したし，ASCIIなどのコード体系の空白を利用して，カタカナを埋め込むことが行なわれた。いわゆる半角カナというのは，この名残と見ることができる（前掲，横山，1982も参照）。

今日，インターネットの先駆者たちの努力によって，日本語によるコミュニケーショ

ン手法は確立し，ユーザサイドではほとんど不自由なく，日本語を読み書きすることができる。少なくともメールやホームページの国内での読み書きの問題が生じることはない。また，今後多言語を一画面に表示するための種々の手法（高橋，2000）によって，国際共通コードを使ったり，その他のエディタを用いたりしても，現在の手法を踏襲して，ユーザが文字の問題にわずらわされることなく，日本語を読み書きできると思われる。

海外との情報のやり取りを考えてみると，すでに指摘したように，共通語としての英語の問題を避けて通るわけにはいかない。海外のホームページを見て情報を得るにしても，逆に自分のホームページを作成して海外に情報を発信するにしても，海外とコミュニケーションをとる手段としては，英語で作成したページを作っておくことが必須であることは間違いない。

個人のホームページではなく，大学や企業のホームページでは，同じページの中に母国語と英語を混在させたり，ボタンを用意して，母国語と英語とを別のページで表示するように切り替えて，情報を取りに来た者に選択させるという方法をとっているものが非常に多い。このようなページでは，英語で表示された部分は更新の頻度が低く，かつ内容も母国語のページよりは簡略化されているケースが多いという調査結果もある。新聞の記事にも，国内の大手企業が作成している英文のホームページで，単語のつづりにミスがあるのが78％もあり，１ヶ月以内の更新がなされていたのは20％にすぎないという，通訳サービスの組織である「ウェブワークス」の調査結果（WebWorks, 2000）が掲載されている（朝日，2000）。

また，ホームページの中には，英語のボタンを押すと，確かに一度は英語で書かれたページが出てくるが，そこに“in Japanese”などという表示があって，結局はその国の言語でしか情報を得られないというものも存在する。私の大学のホームページでも，残念ながら一部はこういう形式である。すなわち，入口には英語の看板がかかっているが，中に入るとその国の言語しか通用しないという，どこの国の実際の店にもあるパターンが，インターネットの中でも起こっているということになる。

こうした状況では，いくらインターネットでは英語が共通語であるといっても，本当に知りたい情報は，その国（ホームページ）の言語に通暁していなければ得られないことになる。実際に，企業などで，必要部品の調達のためにインターネットを検索し，韓国に適切と思われる企業を発見したものの，ハングルの壁に阻まれてその企業とうまくコンタクトが取れなかった例などが報告されている。しかしながら，有力な情報を得るために，お互いに英語での詳しい情報を強いるのは，結局双方の負担増につながり，情報受発信の障害になる。これは，英語のホームページと同じ内容の日本語のホームページがあったときに，どちらを選択するかを考えてみても明らかであろう。

英語でもコミュニケーションが大変（あるいはわずらわしい）のに，英語以外の言語での情報受発信はほとんど不可能と，一般的には思われるが，それをある程度軽減する手段が，以下に述べる機械翻訳の利用である。機械翻訳の技術は，バブル期にやや誇張して宣伝されたために，その反動で逆に不当に低く評価されている面があるが，未熟な面を理解して使えば，ある程度は使える技術である。すなわち，分野や語彙，言いまわし等をある程度限定して使用すれば，完全ではないにしても，ある程度情報を伝えることは可能である。情報発信に対しては，母国語を相手言語に変換し，情報受信に際しては，相手言語を母国語に変換する機械翻訳エンジンを用いることによって，コミュニケーションを円滑化することができる。しかも機械翻訳システムは，最近，ネットワークに対応したものに変化してきている（後述）。以下では，機械翻訳システムを情報受発信

にどのように利用できるかという観点について考察する。

▶ 3　機械翻訳システムとそのネットワークへの対応

図1に機械翻訳システムの現在の様相を示す（横山，1999,2000a）。図の左上の「既存の機械翻訳システム」とは，従来から使われてきた機械翻訳システムのことで，10年くらい前には，大型機やワークステーション上のシステム，あるいは専用機として供給されていたが，今日ではほとんどがパソコン上のシステムとして用いられている。パソコンでは，組み込みソフトとして供給されているものも多い。10年前に比べると，辞書は大規模で詳細になり，専門分野辞書なども充実し，なおかつワープロ等のソフトと互換性のあるものがほとんどである。

図の右半分の円は，ネットワークを示す。この上に２種類の異なる形態のシステムが作られている。一つは，図の右上にあるシステムである。これは，形としては既存の機械翻訳システムと同じであるが，ユーザから見てネットワークの向こう側にあることが，左側のシステムと異なる点である。すなわち，システムの管理は，オンラインで翻訳サービスを提供する会社が行ない，ユーザは，ネットワークを通じて翻訳サービスを受けるというものである。人手による前処理や後処理を含むもの（サービス時間に制限があるものが多い）と，完全自動化されていて24時間サービスを行なうが，機械翻訳システムの入出力のみを用いるものがある。このシステムの特徴は，version upに即応してシステムが更新され，常に最新のシステムが供給されることと，ユーザが提供したデータや情報に基づいて，辞書が頻繁に更新されることである。多数のユーザから寄せられた新しい語を積極的に取り入れる（どのような語を辞書に入れるかの判断はサービス側で行なう）ことによって，最新の語彙がシステム上で供給されるのが左側のシステムとは違う利点である。

これらのシステムは，どちらかというと，機械翻訳を文書処理的な観点で用いているが，最近ではWeb上のブラウザに対応した（いわゆるネットワークサーフィン型の）機械翻訳システムが登場してきた。この種の機械翻訳システムが最初に発売されたのは1995年であった（村田，2000）が，現在では多くのシステムが発売されている。これは，HTMLなどで書かれたホームページを，言語の制御部分とテキスト部分とを区別して，

図１　ネットワーク時代の機械翻訳システム（横山，1999, 2000a）

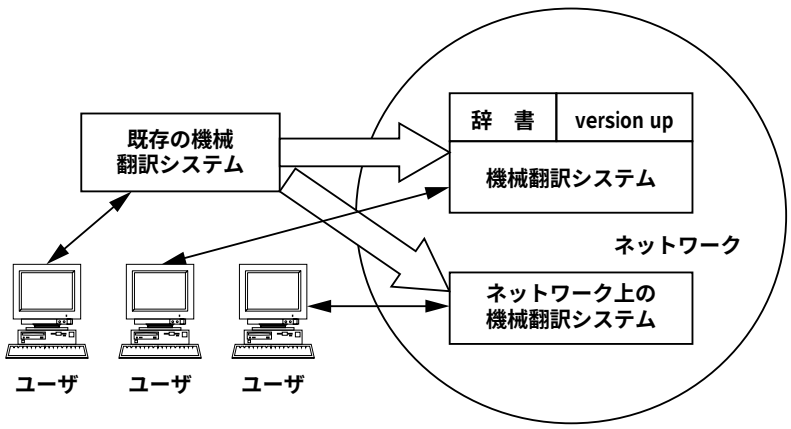


Figure & Table

レイアウトを損なわない形で機械翻訳して提示するもので，Web上のボタンなどを目標言語に翻訳する。その際に，画像データなどを表示するための時間差を利用して機械翻訳するなど，いろいろな工夫がこらされている。村田によれば，これらには，細かく分類すると，通信路に翻訳システムを置くproxy方式，ブラウザが表示しているWebページのデータを取得してその翻訳結果を再度ブラウザに表示させるブラウザ連係方式，図の右上のシステムと類似しているが，Webサイトで翻訳サービスを行なうサーバ方式の３種類がある。最初のタイプのものが，パソコンにのる形で最もよく用いられているが，上にも少し述べたように，翻訳機能や表示機能にいろいろな差がある。これらについては，本紀要の別の論文（宮澤，2001。またMiyazawa, 1999 も参照）に詳しい。第３のタイプには，Alta Vistaなど，検索エンジンに機械翻訳システムを組み込んでおいて，検索をかけたホームページを，所望の言語（英語，フランス語，ドイツ語，スペイン語など。ここでは，1960年代に開発された機械翻訳システムを改良して用いている。日本語サービスは，2000年現在ではまだ行なわれていない）に変換するシステムも含まれる。中には，自分の言語で検索をかけると，その項目を自動的に機械翻訳して，多言語検索（cross-language retrieval）してくれるシステムもある。

上記のようなシステムを利用することによって，情報を受発信する際に障害となる言語の壁の問題を，ある程度取り除くことができる。しかしながら，そのような翻訳が的確に行なわれているかどうかをどう評価するかという重要な問題が起こる。次節では，この点について少し議論する。

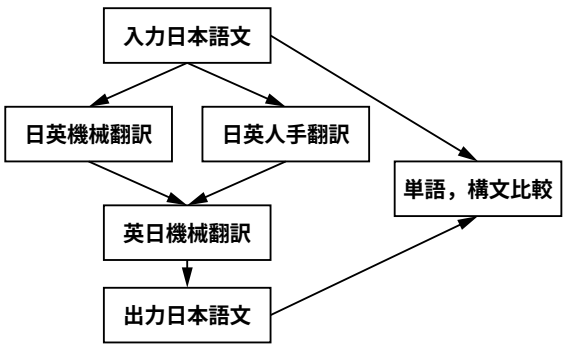
▶ 4　翻訳の評価

機械翻訳では，正しく翻訳されているか，すなわち翻訳結果をどう評価するかということが最も問題となる。特に，自分にとって未知の（あるいは余り得手でない）言語との間の翻訳結果は，それを見たときに正しいかどうかを判断できないという問題がある。情報を受信する場合には，それでもまだ，いくつかの情報（画像など）を手がかりにして，正確さを推定することができるかもしれないが，発信する場合には非常に難しい。

根本的な解決にはならないが，このような場合に用いることのできる評価方法として，「循環翻訳」とか「２方向翻訳」といわれる評価方法がある（Yokoyama, 1999）。以下にこの方法の概要を述べる。

図２に日英翻訳に対する評価方法の概要を示す。考え方としては，「日本語を英語に翻

図２　機械翻訳の評価手順（Yokoyama（1999）を日本語に直したもの）



訳したあとで，もう一度日本語に翻訳し戻してやったときに，最初の日本語と最後の日本語とが類似していれば，その翻訳はおおむね正しい」というものである。まず，入力となる日本語文を，人手（一人）と，機械翻訳システム（この論文では５システム）で英語に直す。日本語文１文に対して６英文の出力が得られる。次に，得られた６英文を，英日機械翻訳システムにかけて（ここでは，５つのシステム。メーカーは日英翻訳の場合とすべてが一致するわけではないし，同じメーカーでも異なるアルゴリズムを用いているので，日英，英日で全く異なるシステムを用いていると考えて差し支えない），日本語文を出力する。この場合には，最終的に全てのシステムで出力が得られたとして，入力日本語文１文に対して，30文の出力日本語文が得られたことになる。この入力と出力を，単語の一致度，係り受けの一致度，並列構造の対応などで評価しようというのが，この評価方法である。

図３　例文とスコア計算（Yokoyama,1999）

＃ITU通信評価基準に基づき各国のディジタル通信量を比較することで，各国のマルチメディアの普及度と経済レベルの相関関係がよく分かる。
(A)19語　(B)8　(C)1
 # By comparing digital communication quantity of each country based on the ITU communication criterion for evaluations, the correlation between prevalence and economy levels of multi-media in each country is well proven.
評価用のITU通信criterionsに基づいた，各国のディジタル通信量の比較によって，各国の多重媒体の普及と経済のレベル間の相関は，上手に証明されている。
(A0)12 (A1)2 (A2)0 (B)8 (C)1
評価のためのITU通信判定基準に基づく個々の国のデジタル通信量を比較することによって，個々の国のマルチメディアの普及と経済レベルの相関がよく証明される。
(A0)13 (A1)1 (A2)2 (B)7 (C)1
% By comparing digital communicative quantity of each country based on ITU communicative evaluation standards, a correlation of a widespread degree of MultiMedia of each country and an economic level is found well.
ITUのコミュニケーション評価規格に基づくデジタルコミュニケーションの量の各国を比較することによって，各国のMultiMediaの広範囲の度と経済レベルの相関関係はよく見つけられる。
(A0)13 (A1)0 (A2)0 (B)5 (C)1
ITU通信評価基準に基づくデジタルの通信量の個々の国を比較することによって，個々の国と経済水準の広範囲に及んだ程度のマルチメディアの相関がよいのが発見される。
(A0)12 (A1)0 (A2)3 (B)2 (C)0

Figure & Table

図３に，いくつかの例文と，スコアの計算を示す。“＃”の後に書かれた日本語が入力文，その後に示したものは，(A)が単語の数，(B)が係り受けの数，(C)が並列構造の数である。次の“＃”の後の英文は，人間が（文脈等を考慮して）翻訳したものである。その後に続く２つの日本語文が，この英文を，日本語文に機械翻訳した結果である。日本語文に続く(A0)は，元の日本語文と完全に一致した単語の数，(A1)は一部が一致した単語の数，(A2)は同義語の数をそれぞれ示す。その次に書かれた“％”の後の英文は，日英機械翻訳システムによって出力された英文である。これを，英日機械翻訳システムを用いて再び日本語文に機械翻訳したものが，それに続く２つの日本語文である。なお，これ

らのA～Cのスコア計算は，この研究では手作業で行なっている。これらのスコアを合わせることによって，ここでは，機械翻訳システムの性能の良し悪しを判断している。

この方法を用いれば，自分の持っている情報を，目標言語で発信したときに，正しい機械翻訳が行なわれているかどうかということをある程度知ることができる。特に，ホームページ上のボタンのように，単語レベルでやり取りすればよいものの場合には，一度目標言語に翻訳したものを戻して，元に復していれば，余り一般的な用語でない限り，かなりの信頼度で正しい翻訳がなされているということができる。

▶ 5　終わりに

現在のホームページの利用の仕方などを見ていると，まだ外国の（特に英語圏の）情報を取ってくるという使い方が多い。しかしながら，今後は，現在国内に向けて発信されている情報を，積極的に海外に向けて発信していく方向に，急速に向かうと考えられる。そのときに情報を発信する側として望みたいのは，外国語で情報を発信するときの補助となるツールが欲しいという要求であろう。ネットワーク上の機械翻訳システムは，この要求をある程度満たすものとして大いに利用すべきである。

機械翻訳システムは，まだ発展途上のシステムである。したがって，現在の形のシステムが販売され始めた当初にあった「万能の機械翻訳システム」といった幻想からは，ほど遠いといえる。そこで，システムの限界をよく知った上で，できるだけ効率的で，人間の能力の限界を補う使用法が望ましい。前節に述べた評価法は，その正確さを測る一つの指標となるものである。しかしながら，まだ翻訳の部分以外を手作業で行なうなど，今後研究すべき課題が多く残されている。我々は，現在アジア太平洋機械翻訳協会（AAMT）のネットワーク翻訳研究会において，これらの問題を解決すべく努力中である。すなわち，評価の多くの部分を機械化して，なるべく客観的な評価ができる方法を現在模索している。

今後もインターネット上での情報受発信の要求はさらに高まり，言語もはるかに多様化していくことが予想される。機械翻訳システムを利用することによって，そのような急速な変化にある程度対応できるのではないかと考えている。

●参 考 文 献

朝日新聞（2000.6.24）「大手企業は英語苦手？」
アジア太平洋機械翻訳協会（AAMT）編（2000）機械翻訳―21世紀のビジョン
岩谷　宏（1995）基礎からわかるインターネット，ちくま新書052，筑摩書房
WebWorks（2000）インターネットによるグローバル情報発信を考える（英文ウェブサイトの現状と未来），WebWorks第１回セミナー資料
鈴木　孝夫（1999）日本人はなぜ英語ができないか，岩波新書新赤版622，岩波書店
高橋　直人，錦見　美貴子，半田　剣一，戸村　哲（2000）Muleを捨てて，Emacsを使おう，情報処理41巻11号，pp.1233-1238
船橋　洋一（2000）あえて英語公用語論，文春新書122，文藝春秋
S. Miyazawa, S. Yokoyama, M. Matsudaira, A. Kumano, S. Kodama, H. Kashioka, Y. Shirokizawa, and Y. Nakajima（1999）Study on Evaluation of WWW MT Systems, Machine Translation Summit VII, pp.290-298
宮澤　信一郎（2001）インターネット翻訳ソフトの現状と将来展望，本紀要
村井　純（1995）インターネット，岩波新書新赤版416，岩波書店
村田　稔樹（2000）発信側翻訳のためのPENSEE for Internet，AAMTセミナー「インターネット翻訳・機械翻訳の展望」予稿集
横山　晶一（1982）OAを支える日本語処理，事務と経営34巻415号，pp.28-32
横山　晶一（1999）ネットワーク翻訳研究会報告，アジア太平洋機械翻訳協会総会資料

S. Yokoyama, A. Kumano, M. Matsudaira, Y. Shirokizawa, M. Kawagoe, S. Kodama, H. Kashioka, T. Ehara, S. Miyazawa, and Y. Nakajima （1999）Qualitative Evaluation of Machine Translation using Two-way MT, Machine Translation Summit VII, pp.568-573
横山　晶一・荻野　孝野（2000a）言語データと言語処理，bit Vol.32, No.2, pp.12-20
横山　晶一（2000b）コンピュータ「弥生時代」における英語独占からネット多言語時代へ，ことばと社会No.3, pp.151-155
横山　晶一（2000c）日本語の発信とネットワーク機械翻訳，日仏円卓会議「日本とその未来―展望とユートピア」予稿集

（横山晶一　山形大学工学部教授）