

機械翻訳(MT)とサーチエンジンを英語学習に活用する

安藤 進

青山学院大学 理工学部機械工学科 講師、テクニカルトランスレーター

機械翻訳(MT)システムと Web のサーチエンジンが学生の英語学習に非常に効果があることを報告する。MT は生徒の心理的な抵抗感を少なくし、生徒の論理的な思考訓練に役立つ。しかし、現在のインターネット対応 MT システムにはかなり問題がある。

アルタビスタが英語の活用辞典として使えることを報告する。ヒット件数と用例を調べることで、自分が直面する特定の問題解決に役立つことがわかった。このやり方を身につけると、研究やビジネスなど多様な面で応用が利く。

Machine translation (MT) and Internet search engines improve your English skills

Susumu Ando

Aoyamagakuin University Mechanical Engineering Lecturer, Technical Translator

This paper suggests that machine translation (MT) systems and Web search engines provide students with powerful tools for learning English. The MT helps students remove their psychological barriers to English and express their ideas more freely than before. Search engines such as AltaVista present the number of hits and example sentences matching the query. The number of hits implies word frequency; example sentences are clues for finding necessary expressions or technical terms based on a specific context.

This paper also evaluates several MT systems currently available for use in instantaneous translation of Web pages by comparing several different MT-translated outputs with my comments for future improvements.

1. はじめに

機械翻訳(MT)システムが商用化されてから10年以上になる。メインフレーム時代に誕生してから、幾多の数奇な運命をたどり、最近インターネットと共に脚光を浴びるようになった。幾度か絶滅の危機に瀕しながら蘇る、まさにソフトウェア業界の七不思議の一つといえるだろう。

システムそのもののデキはともかくとして、学生の英語学習という面で思いがけない効用があることに気がついた。ひとつは心理的な抵抗感をなくしてくれること。もうひとつは言語に対して工学的なアプローチができること。また、人間では思い付かない奇抜な解釈があり得ることも教えてくれる。

サーチエンジンとは奇妙な呼称だが、特定のキーワードをWebページのフィールドに入力することで、Webページやニュースグループの文字データを検索してくれるサイトのことである。例えば、アルタビスタ(<http://www.altavista.digital.com/>)、ホットボット(<http://www.hotbot.com/>)、ヤフー(<http://www.yahoo.com/>)、グー(<http://www.goo.ne.jp>)などだ。

特にヒット件数を調べることで、特定の語句の使用頻度をある程度推定することが可能だ。これらのサーチエンジンを英語の生きた活用辞典として使えることを報告したい。

最初に、事例を紹介しながら、MTとサーチエンジンの利用法について検討する。次に、インターネット対応というたい文句のMTシステムの実力について検討し、今後の課題を提起する。

2. 課題文をMTに翻訳させる

一般に理工系の学生は英語に苦手意識をもつ。教室で課題文を生徒に割り当て、黒板に試訳を書かせる伝統的なやり方を心理的に嫌がる学生が多い。教室で私語をする学生が多いが、

課題について発言する学生は非常に少ない。

まな板に載せる材料をMTにやらせるのがミソだ。これは生徒たちに心理的な抵抗感をなくす効果がある。MTが原文をどう解析し、どう訳文を生成するのか、簡単なステップを紹介しながら、説明すると、生徒も興味を示す。その過程で、構文解析や訳語選択の適否について、それとなくヒントを与えておく。

教室で手を上げて質問する生徒が少ないのは問題意識が少ないからではない。授業をコンピュータ室で実施することで、生徒に各自のコメントや試訳を電子メールでレポートさせると、そのことがわかる。

2.1. 事例の紹介

最近、私の授業で採用した課題文と生徒の反応を紹介する。

2.1.1. 事例1

[課題文1] The absence of separate induction and exhaust strokes in a two-stroke engine is the fundamental difference with four-stroke engines.

[MT訳] 2ストロークエンジンの別々の誘導と排気ストロークの欠如は4サイクルのエンジンをもった基本的な違いである。

[注] 「工業英語3級試験問題」(1996年11月実施)。MTは富士通社のパソコン版Atlas。

私はこの訳例の中から次の個所を取り上げて、どうしてこのような相違がでるのだろうかかと疑問を投げかけた。

two-stroke engine 「2ストロークエンジン」
four-stroke engines 「4サイクルのエンジン」

まず、MTシステムの辞書構成について簡単に説明してから、専門用語辞書にfour-stroke = 「4サイクル」という訳語が記載されている可能性が高いという推定を述べた。

次に、英語と日本語の世界で「サイクル」と「ストローク」のどちらが多く使われているのか調べてみようとして提案した。コンピュータ室でサーチエンジンを使って実習をし、その結果を電子メールでレポートさせた。その結果、次の事実がわかった。

グーによる検索結果 (<http://www.goo.ne.jp>)

4 ストロークエンジン :	25 件
4 サイクルエンジン :	90 件
4 ストロークサイクルエンジン :	0 件
2 ストロークエンジン :	18 件
2 サイクルエンジン :	91 件
2 ストロークサイクルエンジン :	0 件

アルタビスタによる検索結果

(<http://www.altavista.digital.com/>)

two-stroke engine :	300 件
two-cycle engine :	100 件
two-stroke-cycle engine :	15 件
four-stroke engine :	200 件
four-cycle engine :	53 件
four-stroke-cycle engine :	13 件
2-stroke engine:	100 件
2-cycle engine :	78 件
2-stroke-cycle engine :	3 件
4-stroke engine :	100 件
4-cycle engine :	100 件
4-stroke-cycle engine :	5 件

生徒のコメント

生徒の生の声を一部紹介する。

こうして調べてみると「ストロークエンジン」と「サイクルエンジン」は、使われ方の違いがはっきり判る例文を見つけられなかったが、ほぼだいたい同じように使われているようである。全体の印象としては「サイクルエンジン」のほうが多く使われているようだ。また、「ストロークサイクルエンジン」という表現はひとつもなかった。

興味深かったのは、英語では"stroke-engine"も"cycle engine"も、同じように多く使われていたが、ヒット件数から見ると日本語とは逆で、"stroke-engine"という言い方のほうが多かった。そして英語の場合では "stroke-cycle engine" という言い方も、数は少なかったが使われていた。

2.1.2. 事例1の考察

わたしは翌週、これらのコメントを集約して、次の話をした。

・日本語では「ストロークエンジン」より「サイクルエンジン」のほうが多そうだ。

・英語では"stroke-engine"も"cycle engine"もあるようだ。

・いずれにしても、ひとつの文の中で混在させるのはよくないので、どちらかに統一したほうがよい。

私は、生徒のひとりが見つけてきたエンジンの画像を紹介しながら、4 ストロークエンジンは吸入ストローク(行程)と排気ストローク(行程)が別々になっているが、2 ストロークエンジンはそうではないことを確認させた。

また、サーチエンジンのヒット数は必ずしも厳密な意味で頻度数と同じとは言えないが、おそよの検討をつけるには役立つこと、例文を丹念に調べることの大切さを強調した。

MT 訳の欠点

MT の欠点を指摘するのは MT をけなすことが目的ではない。同じ生徒の試訳の場合は心理的な抵抗感があるが、MT という非人間的な機械が出力したものであるので、比較的自由に意見が出しやすい。わたしは次のような MT の問題点を指摘して、試訳を提示した。

・ and のスコープの判定

induction and exhaust strokes は induction (strokes) and exhaust strokes と補って解釈する必要がある。

・ with の解釈

これは非常に難しいが、「～の場合との」といった意味だろう。

・ The absence of separate A and B の解釈

Web 検索で見つけた画像により背景知識が得られたことを踏まえ、「A と B が別々でないこと」と解釈すれば文意に合うと説明した。

この検定試験問題の配点は 15 点。これを受験生の答案として採点すれば、約 6 点に相当するだろう。このような話をすると、課題文に対する取り組み方と目の色が違ってくる。以上をまとめて、次の試訳を披露した。

[試訳] 2 サイクルエンジンは吸入行程と排気行程が別々になっておらず、これが 4 サイクルエンジンとの基本的な相違である。

2.1.3. 事例 2

[課題文] Research has shown that letters from 8 points to 12 points high are easiest to read.

[MT 訳] 研究は、8 個のポイントから 12 ポイント高値へのレターが最も読みやすいのを示した。

[注]「工業英語 3 級試験問題」(1996 年 11 月実施)。MT は富士通社のパソコン版 Atlas。

コンピュータ室で生徒(約 40 名)に「簡単なコメントを付けて自分なりの試訳を提示するように」と指示した。必要に応じて次のような辞書を紹介した。研究社の辞書

(<http://www.aix.or.jp/cgi-bin/cgi-readers>)や E-J/J-E 辞書

(<http://anansi.panix.com/userdirs/tn/j-search.html>)、日英辞書

(<http://www.wg.omron.co.jp/cgi-bin/j-e/dict>)などだ。残念ながら、研究社の辞書は 6 月半ばから有料になり、授業では使えなくなってしまう。

生徒の試訳の中から代表的なものを以下に

示す。

1. 調査した結果、8 個のポイントから 12 個のポイントの高いほうがより読みやすいとわかった。
2. 研究の結果、8 文字から 12 文字位が読みやすいということがわかった。
3. 研究は、8 行から 12 行ぐらいの文章が最も読みやすいのを示した。
4. 研究の結果、8 単語から 12 単語の文章が最も読み易い。
5. 研究論文は、8 個から 12 個の点を文章にまとめたものが最も読みやすい。
6. 研究の結果、8 から 12 以上の点に分かれた手紙が最も読みやすいことを見せた。
7. 研究の結果、8 ポイントから 12 ポイントの大きさの文字が最も読みやすいことがわかった。

生徒からのコメントの一部を紹介する。

・「8 個のポイントから 12 ポイント高値へのレターが」のところが 変に感じる。high のかかりかたがおかしいと思った。しかし、あまり意味が良く分からないのでどのように直したらいいのかわからなかった。

・「8 個のポイントから 12 ポイント高値へのレター」の意味がよくわからない。letter の意味を調べてみたら、一般的に知られている「手紙」の意味のほかに「字(数)」の意味もあることから、「8 字から 12 字以上の単語が一番読みやすい」という感じになると思うが、自信はない。もうお手上げだ。

・「12 ポイント高値」と「レター」という表現が変な気がする。訳の意味が分からなかったのので、どう訳せばいいかは分からないが、変だと思う。

2.1.4. 事例 2 の考察

翌週、この 7 通りの訳例を生徒に示して人気投票をしてみた。「原文を見ないでどれがわかりやすいか」と判断基準を示した。その結果は次の通り。

- (1) 3% (2) 41% (3) 35% (4) 12% (5) 0%
(6) 3% (7) 6%

MT 訳がわかりにくいことはわかるがどうすればよいのかで苦しんだことがわかる。実にバリエーション豊かな反応だ。人気投票をした結果も見事にばらついている。これは人間の主観的な判断がいかに多様性があるかを示している。

MT 訳の問題点はほぼつかめている。point の解釈で「文字」「行」「単語」「要点」など、いずれも「読みやすさ」という点ではよいが、「正確さ」という点では問題だ。point 自体にそのような語義がないからだ。

そこで、ワープロソフトの Word を起動させて、フォントサイズのフィールドを示し、そこに表示されている数字がポイント数であることを指摘した。

これを踏まえて「文字の大きさ」を意味する「ポイント」と解釈し、数字は「個数」「数量」ではなく「番号」と解釈すれば、文意に合うと説明した。したがって、(7)の訳例が正解。これに対する投票数が少なかったのは、文字とポイント数に関する背景知識がないからだという、なるほどとうなずいた。

最後に、次のように要点をまとめた。

- ・「正確さ」と「わかりやすさ」は相反するものだ。
- ・正確さを求めすぎると、語順や訳語選択のわずかなミスで、意味不明になりやすい
- ・わかりやすさを求めすぎると、架空の物語(ウソ)になりやすい。
- ・背景知識と語学力の二刀流で迫る必要がある。

3. 機械翻訳の実力は？

最近、インターネットのホームページを即座に翻訳する機械翻訳システムが発売されている。システムによってはあらかじめ本体に内蔵されているものもある。

先ほど紹介した「アルタビスタ」のヘルプページを試してみた。例文、訳例、コメント、

試訳の順で紹介する。

3.1. 訳例の紹介

このページの中からサーチエンジンを使うために必要な文をいくつか取り上げてみる。

1. Lower-case search will find matches of capitalized words also. For example, paris will find matches for paris, Paris, and PARIS.

[MT 訳例 1]

小文字検索が同じく大文字化された言葉のマッチを見いだすでしょう。例えば、paris が paris、パリとパリのためにマッチを見いだすでしょう。

[MT 訳例 2]

小文字の探索は、さらに大文字化された言葉のマッチを見つけるでしょう。例えば、paris は、paris、パリおよび PARIS に匹敵する人を見つけるでしょう。

[MT 訳例 3]

低いほうのケース調査は、また資本化された言葉のマッチを見つけるだろう。たとえば、パリは、パリ、パリ、および PARIS にマッチを見つけるだろう。

[コメント]

翻訳対象の判定と訳語選択に誤りがある。paris, Paris, and PARIS は具体的な入力例なので、翻訳対象外にすべきだ。

Lower-case search は「小文字で入力して検索する」。「低いほうのケース調査」(訳例 3)は意味不明。

capitalized words は「大文字の単語」。「資本化された言葉」(訳例 3)は意味不明。

matches は「一致した語句」。「PARIS に匹敵する人」(訳例 2)は意味不明。

[試訳]

小文字で単語を入力すると、大文字の単語もいっしょに検索される。例えば、paris と入力す

ると、paris も Paris も PARIS も検索対象になる。

2. Capital letters in a search will force an exact case match on the entire word. For example, submitting a query for parIS will search only for matches of parIS. (Don't be surprised if there are none.)

[MT 訳例 1]

検索での首都手紙が全部の言葉の上に正確なケース試合を強制するでしょう。例えば、parIS のために質問を提出することは parIS のマッチをだけ捜すでしょう。|| (|| そうしない || あるかどうかに驚いている || なし。||)

[MT 訳例 2]

探索中の頭文字は全単語を正確なケースマッチに強要するだろう。例えば、parIS のための質問の提出は、parIS のマッチをのみ探索するでしょう。(1 つもない場合、驚かない。)

[MT 訳例 3]

調査における大文字は、言葉全体について正確なケース・マッチを強要するだろう。たとえば、パリに質問を提出することは、単にパリのマッチを捜すだろう。(★★ - - - しないこと - - - いずれもないならば、驚く。)

[コメント]

force A on B は「B に A を強制的に押し付ける」。この解析はどれも失敗。

Capital letters は「大文字」。

[訳例 1] の「首都手紙」は、製品版では「大文字」になった。「頭文字」(訳例 2) は、厳密には正しくない。

an exact case match は「大文字と小文字を区別として完全に一致した (単語を検索する)」という意味。

[試訳]

大文字を含んだ単語を入力すると、大文字と小

文字を区別して完全に一致した単語だけが検索される。例えば、parIS と入力すると、parIS だけが検索対象になる。(検索結果がゼロ件でも驚かないでください。)

3. Precede a required word or phrase with + and a prohibited one with -.

[MT 訳例 1]

|| + と禁止された (の・もの・人) で必要とされる言葉あるいは句より先に起こる || 一緒に || - ||。

[MT 訳例 2]

必修の単語、+ を備えた句および - を備えた禁止された一つに先行してください。

[MT 訳例 3]

+ と禁じられたものを必要な言葉またはフレーズの前置きとすること。

[コメント]

Precede A with B は「A の前に B を付ける」。この構文が正しく解析できた例はなかった。and のスコープの判定に誤りがあります。and precede と補って解析する必要がある。

[試訳]

検索対象に含めるかどうかの指定ができる。語句の先頭に + 記号を付けると、その語句は検索対象になる。 - 記号を付けると、その語句は検索対象から除外される。

3.2. MT システムの考察

ここで紹介した機械翻訳例は、体験版として無料で配布されているもの(1996 年 10 月時点)から採用した。どのシステムが優れているかを評価するのは、ここでの主題ではないが、少なくともここで紹介した例に限って言えば、実用に役立つとはとてもいえない。

インターネット利用を標榜している以上、代表的なホームページによく出てくる表現や語

句については、もう少しましな翻訳が欲しい。
特に、サーチエンジンはこれから利用頻度が高
まることが予想されるので、対策が望まれる。

4. まとめ

MT と Web サーチエンジンを利用した事例
を紹介した。MT の利点は次のようにまとめら
れる。

- ・生徒の心理的な抵抗感をなくせる。
- ・論理的な思考訓練に役立つ。
- ・思いがけない解釈が成立することを知る。

ここでは紹介できなかったが、この検定試験
問題の和訳 8 問と英訳 5 問題を MT に翻訳さ
せた結果は、和訳が 45%、英訳が 65%のでき
だった。実際に受験すれば、和訳は不合格だが、
英訳は合格すれすれだろう。特に、英訳では用
語を正しくつづける生徒が非常に少ないのに対
して、MT はおおむねよくできていたことが印
象的だ。

しかし、別の MT システムで Web テキスト
を翻訳させた結果はひどいものだった。これで
はほとんど実用にならない。対象文書の相違が
原因だとしても、これでインターネット対応は
苦しい。

Web のテキストデータを収集し、それを頻
度順にソートして基礎データとし、Web で頻
出する動詞のパターンや表現例を研究する必要
がある。今後の課題は Web の語彙調査と表現
の研究だと思う。

サーチエンジンの利点は次のとおり。

- ・アルタビスタが英語の活用辞典として使える。
- ・ヒット件数と用例がポイントになる。

数字の妥当性については、検索された用例
を調べることを検証できる。実際に用例の URL
をクリックすると、Not Found(見つかりませ
ん)というメッセージが表示されることも珍し
くはない。その意味では不正確な統計データで

あるが、英語学習に利用する目的ならこの程度
の誤差は許容できる。

最近、経済や化学などそれぞれの分野ごと
に「活用辞典」が各種出版されるようになった
が、個別の文や表現語句について調べたいとき
に、なかなか該当する表現が見つからないこと
が多い。サーチエンジンは、それぞれの個別の
文脈に合った使い方が可能である。

なお、拙著「インターネット英語の読み方
&書き方&調べ方」(共立出版、1997年3月10
日発行、1,957円、ISBN4-320-07139-5 C3050
P3502)を参照されたい。

連絡先: sando@twics.com