

タイトル:用語集超活用ソフト**SuperHT<sup>3</sup>**のご紹介

著者:有限会社アトリエ・ワン **SuperHT<sup>3</sup>**事業室 貝島良太

本文:

私は、昨年の4月1日 AAMT の個人会員になりました。

個人会員としては新参者ですが、株式会社日立国際ビジネスに在職しておりましたとき、早い段階からJAMTの法人会員になっておりました。同社(当初の社名は株式会社国際ビジネスサービス)で、私は翻訳部門を担当しておりました。その後開発部門を発足させ、用語集超活用ソフト**HT<sup>3</sup>**(エイチティースリー)と改良版の**SuperHT<sup>3</sup>**(スーパーエイチティースリー)を開発しました。昨年6月に同社を退職し、SOHOで引き続き同ソフトに関する業務を行っております。このたび事務局の方から、機械翻訳に関連する内容での投稿依頼を頂きましたので、私と機械翻訳の馴初めと、私が開発いたしましたソフトについて述べさせていただきます。

ちなみにAAMT Journal No.19 (June 1997)の新製品紹介欄に**HT<sup>3</sup>**について掲載させていただいております。

### 私と機械翻訳

私が機械翻訳に初めて触れたのは1984年1月のことです。当時私は株式会社日立製作所の輸出営業所コンピューター部に在籍し、欧州向け小型ディスク販売を担当しておりました。新年早々に事業部長から突然呼ばれ、「明日からシ研(川崎市にある日立製作所のシステム開発研究所のこと)に行ってもら。目的は、現在シ研で開発中の日英機械翻訳の開発を手伝うこと。2年以内に立ち上げて、機械翻訳を事業の柱とする翻訳会社を設立したい」と言われました。

それまでは私は機械翻訳のことをSF小説や鉄腕アトムの世界のことと思っていました。いわば外野席から無責任に見ていましたし、第一専門知識がゼロですから、与えられた任務をまったく遂行できない、と先ずは辞退いたしましたが、結局は宮仕えの悲しさ、「開発を横から見て、どの程度使えるかの評価や、改良の提案をするればよい」とのことでしたので引き受けました。それから2年間、右も左も分からない状態で研究所通いが始まりました。

当時、パソコンなどはもちろんありません。研究所の計算機センターにあるホストコンピューター(HITAC M200Hという何億円もする汎用コンピューターで、空調された部屋にでーんと据え付けられていました)に研究室にある端末(T-560/20)から、日本語を入力しておく、翌朝計算機センターに翻訳結果がプリントアウトされており、それを取ってきて、やおら英語の出来を評価するというものでした。プリントアウトを取りに行くのを忘れていたりすると、「早くとりに来るように」とのお叱りの電話が計算機センターからあったことも何度かあります。

この2年間で、日立以外の機械翻訳のデモを見学したり、各社の開発者との意見交換などかなり精力的な活動をいたしました。スタンフォード大学で開催されたCOLING'84にも参加しました。テキサス大学の言語研究所に、英語の冠詞と不定冠詞の使い分け規則を作ってもらったり、New YorkのSmart社でPEP(Plain English Program:平易な技術英文を書くソフト)の導入検討もしました。いずれも英文生成に役立つためです。という訳で、この2年間で相当機械翻訳の知識が得られたことは確かです。

蛇足ですが、当時の日立の日英機械翻訳のニックネームはATHENE/N(ギリシャ神話の学問の女神)でしたが、アテネは(翻訳結果が)「合ってねー」に通じ縁起が悪いからと、HICATS JE(ハイキッツジェイイー)という猫のような名前になりました。

### 機械翻訳の問題点

翻訳の質を上げるために文法規則や辞書の整備をしなければなりません。その規則と辞書を作るた

めの例文に対しては、当然良い結果が出るようになります。しかし、人間にとって例文と同程度と思われる文章でも機械から見ると別物で、でたらめな訳文が出てしまい予想が裏切られることがしばしばあります。これを回避するために更に規則を追加するのですが、次第に規則同士の矛盾が生じて自滅する確率が高くなってしまいます。また、理論的に幾通りもの翻訳結果が出るのがほとんどで、それを見ると同じような文章が場合によっては100文くらい羅列され、その中から正しそうなものを探すだけでも大変な作業でした。自分で翻訳したほうがよほど早いと感じたことがたくさんあります。少なくともまったく同一の訳文は一つだけ表示するという改善提案をし、取り入れてもらったことが懐かしいです。

そういう訳で当時の機械翻訳はエンドレスな開発となり、問題は山積でした。これらの見直し、対策やら改良で「使用に耐えるものは後2年間待って欲しい」と常々研究者から言われていました。そのころは私もうぶでしたから、「そういうものか」と信じておりましたが、2年たっても良い結果には恵まれませんでした。

シ研での2年の刑期(!)を終えるにあたり「機械翻訳の翻訳事業への導入」について社内報告書をまとめました。結論は、「機械翻訳を翻訳事業に導入するのは時期尚早。経済性を無視すれば可能」というものです。当時の試算で、機械翻訳は手翻訳(通常の翻訳者による翻訳のこと)の場合と比べ、2倍程度の費用がかかると結論付けました。手書き原稿からのテキスト入力費用、プレエディット費用、CPU 使用料、ポストエディット費用、更に高額なハードとソフトのリース料を見込むからです。特に、当時の汎用コンピュータの使用料は割高で1文章訳すのに10円から50円程度かかっていたと思います。とくに、始末が悪いのは、機械にとって難解なものは相対的にCUP時間が長くなり、その挙句最後に日本語交じりの、訳文とはいえないものが出てくるのです。失敗したもののほど金がかかるといふ恐ろしいものでした。

1987年に新会社は設立され、私は新会社の翻訳部門に移ることになりました。結局その報告書が効いて(?)機械翻訳は導入しないで普通の翻訳業務を含むドキュメンテーション事業を立ち上げました。1991年 JAMT の発足に当たり、いつ実用に供する機械翻訳が出てきても良いように最新情報を得るために法人会員になりました。

機械翻訳に携わる多くの方々に完成時期をお伺いすると、口をそろえたように「あと2年もあれば何とかなる」と言われました。あれから20年経ったのですが、やはり「あと2年」はかかるだろうと思います。やはり、最大の問題点は自由入力の自然言語に対応できないことです。これはあまりにもハードルが高すぎて、無理であるということだと思います。制限文法で書かれた原文であれば、まあまあ結果が出ますが、それですと素人では使えず実用にはなりません。自由入力文を制限文法で書き直すのに手間(お金)がかかりますし、書き直したら必ず良い結果が出るかはやってみないと分からないという点は永久課題として残ります。現在は、コンピューター環境が20年前と比べて隔世の感がありますから、自分のパソコンで手軽な機械翻訳ソフトを遊んでみる事ができます。入力のやすさなど多くの進化は十分感じますが、翻訳性能自体は残念ながら20年前の研究所レベルと大差ないように感じます。

次に大きな問題点は、訳文を読んでみないと正確さがわからないという点です。ということは、日英/英日のように英語の知識が多少でもある場合は、自分で原文と訳文を見て、あっているとかいないとかの評価ができますが、まったく知らない言語と日本語の場合は訳文が正しいのか判断がつかないという点です。マーズ・アタックという火星人が地球に攻めてくる SF コメディー映画があります。一人の科学者が機械翻訳装置を作り英語と火星語を翻訳するところがあります。人間からの最初の歓迎「ようこそ」の翻訳結果を聞いた火星人たちが突然激怒し、その場にいた人々を皆殺しにしたのをきっかけに、人間と火星人の間で戦争が始まるというものです。結局は人間が勝つのですが、翻訳結果

が人間には分からないので、なぜ火星人が怒ったのかが分からず、人間も火星人のことを失礼な連中だということになったのです。本来の機械翻訳はまったく相手の言語が分からない状態で安全に使えなくてはならないはずで、現状の日英/英日機械翻訳を使って、あつてるとかいけないとか言っているようでは機械翻訳の先はまだまだ遠いと言わざるを得ません。

### こんなソフトはないものか

前職での翻訳業務は、マニュアルの日英/英日翻訳が業務の大半を占めておりました。多くの場合顧客からは原稿と共に用語集が支給されます。当然顧客は用語集にある訳語は翻訳文に使われていることを期待します。しかし、翻訳作業には当該分野に強い翻訳者を当てますから、翻訳者自身が用語をよく知っており、支給された用語集をあまり参照しないで翻訳をすることになります。そうすると、顧客指定の用語集にある訳語があまり使われない翻訳が出来ることも多く、顧客からのクレームを受けることになります。

翻訳の仕事の多くは短納期を要求されます。もし、このような不具合が見つかりますと、顧客からは「明日の朝までチェックして再提出するように」と言われます。このようなクレームはほとんどが夕方退勤しようとする直前に電話で来るのです。こうなると悲惨です。翻訳者に再チェックを依頼しても、「別の仕事を始めているので直ぐにはそのチェックに手がつけられない」、「用語集の訳語を全部使うなんていう作業は、到底できない相談」で、最後は「この納期と料金では無理」、「いや、発注のとき用語集の用語を使うように指示したはず」、「重要と思われるところは用語集の用語を使ったのだから良いではないか」、「もう貴社とは仕事ができない」などのお定まりの応酬の末、けんかになります。顧客には内輪のことはあまり言えませんので、間に入ったコーディネーターが用語集と訳文を徹夜でチェックするようなことがしばしばありました。それでもすべてを潰せるものではありません。

現状の機械翻訳は業務用にはまだ使えない。だから翻訳者に翻訳してもらわなければならない。しかし、用語集の訳語を思うようには使ってもらえない。用語集の用語が訳語にきちんと使用されたかの検収にも多大な時間がかかる。さてどうするか。

そこで考案したのが、

翻訳原稿に訳語付与を完全に正確に行うこと。(翻訳者はそれを利用して翻訳する)

翻訳結果の用語確認ができること。

日英翻訳のために作った用語集を英日でもリバーシブルに使えること。

両言語とも異表記にも対応できること。

用語集(辞書)はExcelで作成出来ること。

複数の用語集を優先順位をつけて使用できること。

翻訳のヒントや解説を参照できること。

ができれば、相当便利なものになるであろうというものでした。当時の既存のソフトをいろいろ探しましたが私のニーズをちょうど満たすものはありませんでした。そこで機械翻訳メーカー各社に、訳語付与までで止める改造をする場合の見積を依頼しましたところ、5千万円から2億円くらいのことを言われ、やむなく用語集超活用ソフトHT<sup>3</sup>の自社開発に踏み切った次第です。

このアイデアのユニーク性が認められ、1999年には日本の、また2000年には米国の特許を取得することが出来ました。

### エンドレスな欲望

HT<sup>3</sup>は1996年6月ごろから開発を開始し、翌年1月に製品化することが出来ました。財団法人日本規格協会とタイアップし、JIS工業用語大辞典第4版対応のCD-ROM with HT<sup>3</sup>を1998年に発売いたしました。HT<sup>3</sup>のユーザーからは、もっと高速で処理したい、書籍版の工業用語大辞典に出ている図表や式を表示できないか、見出し語のボールド/イタリック/上付き文字/下付き文字の文字属性も

文字列チェックと訳語付与の対象にして欲しい、辞書に登録出来る用語の重複を可能にできないか、解説に表示された文章をコピーして使いたい、見出し語の最長60バイトは短すぎる、等々多くの要望が寄せられました。これらはいわゆる改良と言う段階では済まされるものではなく、完全にプログラムの書き直しをしなくては対応できないことが分かりました。一大決心をしてこれらユーザーの要求をすべて満たす新しいプログラムを開発し、**SuperHT<sup>3</sup>**と命名、2001年3月に発売開始いたしました。日本規格協会からはJIS工業用語大辞典第5版対応のCD-ROM with **SuperHT<sup>3</sup>**の発注を受け2001年9月に発売開始いたし現在に至っております。その後も、ユーザーからの要求は続いております。最近では、ヒットした文字列に色をつけるのではなく、MS-Wordの蛍光ペンを使ってヒットを表示することで、原稿にすでに付いている文字属性を変えないで欲しいという改良要求にも対応しております。

ソフトを開発してしみじみと実感することですが、人間の欲望はとどまることを知らないと言うことです。正直怖いです。

### **SuperHT<sup>3</sup>はこんなソフトです**

訳語付与による翻訳支援に加え、最近**SuperHT<sup>3</sup>**がマニュアルなどの日本語文書の用字表現の統一にも利用され始めました。もともとこのソフトは、AAMTの会員各社が開発されている(原文の構文解析をしてから訳文を生成するという)本格的なものではありません。辞書との整合性をチェックしますから、どちらかという翻訳メモリーソフトに近いソフトですが、それとも違う独自のものです。

翻訳目的には**SuperHT<sup>3</sup>**用語集(辞書)に、2カ国語でペアになる見出し語(標準表記)と、各言語で標準表記に対する異表記を必要な数だけ登録できます。用語の標準表記のペアがかにの甲羅、各言語の異表記部分が手足、しかもこの辞書構造が左右どちらからでも利用できるよになっていることから、「かにの原理」と私は呼んでいます。異表記入力は必須ではありませんが、入力しておくことヒットの可能性が高くなり、その分便利に使えます。

**SuperHT<sup>3</sup>**と、通常の機械翻訳ソフトや翻訳メモリーソフトとの最大の違いは、翻訳したあとの用語のチェックが出来るという点です。通常は、機械翻訳でも手翻訳でも翻訳文の用語チェックは人手でせざるを得ませんが、**SuperHT<sup>3</sup>**の用語確認機能を使えば、用語集にある標準表記と異表記を翻訳結果の文書中で色分け表示出来ます。このような使い方を想定してこのソフトを開発したのですが、最近では翻訳者が英日翻訳する場合や、日本語マニュアルを執筆する場合に、用字と表現の統一用に**SuperHT<sup>3</sup>**の日本語部分だけ(すなわち「かに」の半身だけ)を利用することが始まっています。例えば、「どうぞお座りください」「キップを下さい」のように「ください/下さい」の正しい使い分けが、辞書の作り方によって出来るようになります。また、常用漢字1945字以外の文字を見つけることも出来るようになりました。

辞書の作り方次第で、部品番号と部品名の置換や付与、ある数字の組み合わせを乱数表の中から見つけることなども出来ます。

以上、随筆状態でとりとめのない内容になってしまいましたけれど、用語集超活用ソフト**SuperHT<sup>3</sup>**に興味を少しでも持っていただければと思います。今後ともよろしく申し上げます。

問合せ先： 有限会社アトリエ・ワン (Atelier Bow-Wow) **SuperHT<sup>3</sup>**事業室 貝島良太  
roy\_kaijima@h8.dion.ne.jp

\* \* \*