



自然言語処理技術による情報マネジメントの実際

機械翻訳システムによる知的活動の推進

企業内業務の効率化を目的に機械翻訳システムの導入を進めてきた。本稿では機械翻訳システム上に存在する問題を明確にし、翻訳者の業務効率化に焦点をあてた課題解決への取り組み、機械翻訳ユーザー層の拡大への施策など、導入／展開の経過に従って述べる。

次に機械翻訳システムを導入することにより、社内にどのような変革が起こったかを述べ、機械翻訳システムの導入を成功させるための要因を示唆し、知的活動推進への貢献の可能性について述べる。

キヤノン（株）情報通信システム本部

畠中 伸敏

hatanaka.nobutoshi@canon.co.jp

（株）東芝 研究開発センター知識メディアラボラトリー

平川 秀樹

hideki.hirakawa@toshiba.co.jp

機械翻訳システム導入の狙い

機械翻訳システムの取り組みは、1989年複写機事業部のサービス部門から始まった。当時の海外交信の手段は手紙とFAXが主流であった。その交信内容はサービス情報、市場対策情報、新製品品質情報などで、いかに速く正確に東京の本社から全世界の拠点やディーラーに、英語に翻訳して伝えるかが課題であった。しかしながら、送付日数、発行量ともに、担当部門の大きな負荷となり、翻訳を含めた海外との交信手番は約10～24日を要した。

また逆に、市場で発生した品質トラブルや製品クレームを、開発部門や生産部門にタイムリーにフィードバックし原因を突き止め市場対応する必要があったが、英語から日本語への翻訳が大きな言語障壁となっていた。

他方、各拠点からの各国言語への翻訳は拠点ごとに翻訳を行っている。ヨーロッパの拠点では、英語からドイツ語、フランス語などの各国言語に翻訳するスタッフがおり、現地での翻訳システムが稼働している。

したがってキヤノンの本社としては、日本語から英語への翻訳が緊急の課題であった。

導入時の課題

- 1) 日英翻訳が翻訳業務の90%を占め、システム化の要求が存在する。
- 2) 海外との交信は即時性が要求される。メールによる交信：10日～24日。
- 3) 海外との交信回数が多く、量的な負荷も大きい。

機械翻訳の取り組みの経過

機械翻訳の導入と課題

1989年、海外との言語障壁を解決するために、機械翻訳システムの評価・導入を図ったが、実運用を行う上では、「使いづらさ」、「翻訳結果の悪さ」が機械翻訳の大きな問題であった。

原文	目標文	機械翻訳結果	判定
ふた	cap (帽子状の)ふた	lid ふた / くちびる	×
補給口	opening	mouth (人間の)口	×
取っ手	tab (先の金具)	knob (ドア, 引き出しの握り)	×
たたき	tap (トナーボトルの底を)たたく	beat 打つ strike (打撃を)加える	×
(ふたを)取り外す (シールを)はがす (トナーカートリッジを1本)補給する	remove (はずす) remove ~ from (取り除く) supply (供給する / 調達する)	detach (分離する) peel off 皮をむく, <樹皮をなどを>はぐ replenish (燃料を)つぐ	×
前ドア	front door	door (前 = front が訳されない) pre door / before ドア	×
(手前に)引き出す (カートリッジの)先端 (ホルダーの)溝 回転する (プラスコロナを)かける 汚れる (ごみではない) 転写不良 裏汚れ 定着器 高抵抗 自動両面コピー 自動多重コピー 分離不良	pull it out tip 先, 先端 (point) groove turn is applied to soil 汚れ, 汚物 poor transfer soiling on the back of the paper fixing assembly high electrical resistance automatic two-sides copy mode automatic overlay copy mode poor separation	draw out (図面を)引く top いただき, 上部 slot 細長い小さい溝 rotate 回転 (循環, 交代) する is put on (乗せる) dirt ごみ, ほこり transfer defect dirt on the reverse side of the copy form setting machine highly resistant 副詞が名詞を修飾 automatic double-sided copying automatic multiple copying separation defect	×
(トナーが)空の	empty	empty / sky	○

表 -1 機械翻訳システムによる単語訳の例 (1989 年当時)

当時の機械翻訳システムは、大型コンピュータもしくは UNIX マシン上で動作したためマシンそのものの保守を必要とし、マシンの立ち上げ、シャット起動時の画面の文字が小さく、画面の操作性が悪く、また、システムダウンやハングが生じた場合は、コマンドベースの操作が必要となった。翻訳業務に従事する人は、コンピュータの知識のない事務職から構成されることから、機械翻訳の使用現場で操作上のトラブルが多発し、機械翻訳システム嫌いを助長する結果となった。問題点の1つは「使いづらい」ことで、苦勞して使いこなしても、それに見合った翻訳性能が得られなかったことである。2つ目は「翻訳結果の悪さ」である。

表 -1 は当時の代表的機械翻訳システムの単語訳の比較を行ったものである。たとえば「補給口」は「opening」と訳すべきところを、動物の口を示す「mouth」と訳され、トナーボトルの底を軽く「叩く」は、「beat (打つ)」, 「strike (打撃を) 加える」などと訳される。この訳ではトナーボトルの底を壊されかねないと、翻訳の専門家からは指摘を受けた。またトナーが「空の」は「empty」が正しい訳であるが、「sky」と訳された。

機械翻訳品質の向上

翻訳品質の第1の問題は、機械翻訳システムが正しい訳語を保有していても、機械翻訳のパース(解析機)が

日本語を文節単位に分割し英語の語順に並び換える段階で、語の結びつきの情報を活用できないことと考えた。

表 -1 に示すように「ドア」に「前」が隣接すると、「前」の訳は「front」で、「pre」「previous」「before」「former」などの語は隣接することができない。

言葉というものは互いに結びつきを持って集まっていることから、語の互いの結びつきの情報を保持しながら原文を解析し、日本語から英語へ変換し、構文生成することを考えた。

同様に名詞句、動詞句、副詞句においても、句には必ずヘッド(主要要素)が存在し、名詞句においては名詞が、動詞句においては動詞が、副詞句においては副詞がヘッドとなっている。これらのヘッド以外の言葉は下位に従属している。たとえば「ドラムを取り付ける」は「attach (取り付ける)」という語をヘッドとして、「ドラム」を目的語としている(下位に従属している)。同様に「複写機を取り付ける」は「install (取り付ける)」をヘッドとして、「複写機」を目的語としている。「ビスを取り付ける」は「fit (取り付ける)」という語をヘッドとして、「ビス」を目的語としている。

このことから、語の連鎖の情報をルールとして登録し条件が合えば発火する仕組みを、機械翻訳システムの機能として追加した。つまり共起ルールを知識ベースとして登録する機能を機械翻訳システムに追加した。

動詞の共起

verb:取り付ける(1)=attach;ドラム/保護/シート/(を);感光/ドラム(に)=(on)
 verb:取り付ける(1)=install;複写機(を)
 verb:取り付ける(1)=fit;ビス(を)
 verb:取り付ける(1)=mount;感光/ドラム(を);複写機(に)

機能表現共起 (副詞, 形容詞)

Adverb:応じて(3)=according to(C2);((f);*倍数(に)=(0));
 (出力する(s)=send;*信号(を)=(0);*基板(に)=(to))
 Adjective:応じて(3)=to suit(C3);((c);*測定値(に)=(0));
 (出力する(s)=send;*信号(を)=(0);*回路(に)=(to))
 Adverb:応じて(2)=in keeping with(B1);(c);*信号(に)=(0);
 (変化する(s)=vary;*出力(は)=(s))

図-1 知識ベースの充実：共起ルールの登録

同時に複写機 NP6650 のサービスマニュアルを分析して、単語辞書の整理 (約 6,000 語)、複写機事業分野固有の文法則の整理を行いルールとして追加した。また図-1のような動詞・副詞の共起ルール (約 3,000 語) を知識ベースとして構築し、辞書として登録を行った。

この結果、たとえば図-2に示すように、「応じて」という機能表現が「to suit」「according to」「based on」と適格に訳し分けされるようになった。このように機械翻訳システムによる翻訳結果の性能を向上させることができ、一般ユーザからはよい評価を得た。しかしながら翻訳性能が十分でないことも含め、プロの翻訳家からは要求水準が高く、受け入れられなかった。

機械翻訳サービスの展開

1990年代を通して継続して機械翻訳の問題点を解決するための施策を実施した。施策の内容は、前節で述べた機械翻訳の問題点に対応して「知識ベースの構築」と「操作性の向上」を同時並行的に進めた。図-3の下部にこの取り組みについて示す。図-3の知識ベースの三角形は取り組みの重点の度合いを示している。

「使いづらい」という課題については GUI の画面が使える Macintosh や NeXT から利用できるシステムとし、操作性を向上させた。また翻訳システムが必要と思われる部署に端末を設置しクライアント・サーバ系でできるようにし、機械翻訳が利用できるサイトを増大させた。プロの翻訳家のいる 4 サイトから 10 サイトでの運用へと増大した。

1990年代中頃からイントラネットが普及し、機械翻訳システムの普及を促進するため、1997年に機械翻訳システムをイントラネット上に載せ、Webブラウザ画面によりアクセスできるようにした。この結果、1日平均 87 名に使用者が増大した。この原因は Web の画面が

目標文

Determines the bias to suit the measurement. Fit a screw. The rollers are rotating during copying/ "1" at time of operation. Alternates between "1" and "0" during M3 rotates. Fit the cable. Adjust the target potential according to the specified condition. The microprocessor based on the reduction ratio, calculates the size of copy paper. Remove the seal from the opening of the paper bottle. Remove the copyboard glass retainer (two screws) and remove the copy board glass. Remove four screws and remove the hopper cover. Check each guide plate for deformation. Change the position of the lens as shown in Figure 3-18. The scanner is moved in reverse by changing the direction. The scanner is moved in reverse by changing the direction. The rotation speed varies depending on the reproduction ratio. After the AE measurement, make a copy. Remove the size index plate (vertical). Detach the drive wire. Remove one screw. Remove three connectors.

原文 (複写機NP6650のサービスマニュアル)

測定値に応じてバイスを決定する。ビスを取り付ける。コピー中はローラは回転する。動作中「1」になる。M3回転中「0」を繰り返す。ワイヤを取り付ける。規格条件に応じて目標電位を調整する。マイクロコンピュータは応じて原稿の寸法を計算する。トナーボトルの補給口に貼ってあるシールをはがす。原稿台ガラス瓶蓋を外し(ビス2本)、原稿台ガラスを取り外す。ビス4本を外し、ホッパーカバーを外す。各ガイド板に裏形がないかをチェックする。図3-18に示すようにレンズの位置を変える。光学系は方向を変えることにより後進を行っている。光学系は回転方向を変えることにより前進を行っている。回転速度は倍率によって変化する。AE自動調整後、コピーする。縦サイズ指標を取り外す。駆動ワイヤを外す。ビス1本を外す。コネクタ3個を外す。

↓ 複写機辞書の構築
共起ルールの登録

機械翻訳結果

Determine the bias to suit the measurement. Fit the screw. The roller rotates during the copying. It goes "1" at time of the operation. "1" and "0" is repeated during the M3 rotates. Fit the cable. Adjust the target potential according to the specified condition. The microprocessor calculates the size of Document based on the reproduction ratio. Remove the seal stuck on the opening of the copy board glass. Remove four screws and remove the hopper cover. Check for the deformation to each paper guide plate. Change the position of the lens as shown in the Figure 3-18. The optical system makes the reverse by changing the direction. The rotation speed varies depending on the reproduction ratio. Make a copy after The AE measurement. Remove the size index plate (vertical). Detach the drive wire. Remove screw. Remove three connectors.

図-2 知識ベースを充実後の翻訳結果

非常に操作性に優れ、単純な機能部分だけに限定したため操作が容易であったことにある。第2の原因は、イントラネットを通して機械翻訳が使えるサイトがパソコンの数だけ増大したことにある。

1999年には、基本辞書 50 万語、専門用語辞書 56 万語の導入を図った。このことは知識ベースの増大に繋がり、翻訳性能を向上させ、使用者に大きな安心感を与えた。図-3に示すように、イントラネットの普及も含めて、大規模辞書の導入を契機に、使用者が1日平均 250 名に増大し、現在では1日平均 350 名となった。

機械翻訳システムのブレイクスルーの要因としては、「Web化」による使いやすいインターフェースの提供と、「辞書の拡充 (基本辞書 50 万語、専門用語辞書 56 万語)」によるベース性能向上であり、以上の経緯を経て、機械翻訳システム導入の成果が得られたと判断する。

利用状況

使用者各人の利用頻度

ここでは使用者各人が機械翻訳システムをどのように使用しているかを利用頻度の観点から調査し分析する。

平成 14 年 7 月 12 日から 7 月 31 日までの実稼働日数 13 日間のアクセスログをもとに、この間に同じ使用者が利用した回数を調査した結果、次のようになった。

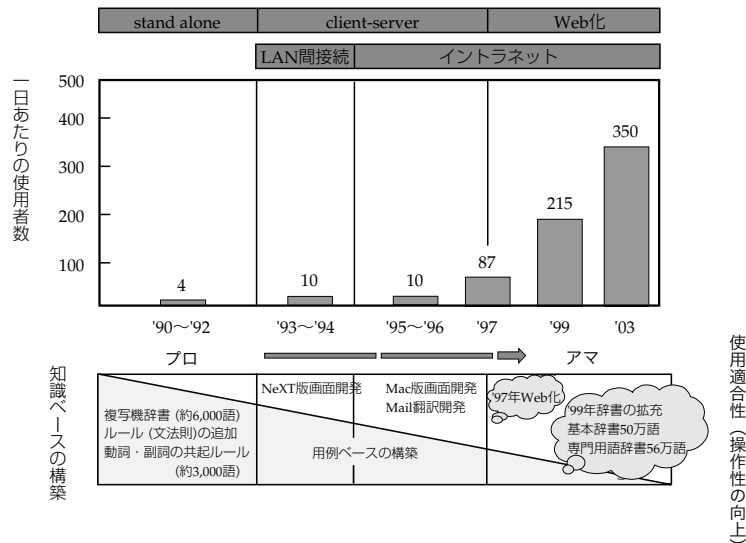


図-3 機械翻訳システムの利用者数の推移

使用回数の内訳

- ・利用頻度 1 回 : 44 名 25%
- ・利用頻度 2 ~ 3 回 : 40 名 23%
- ・利用頻度 4 回以上 : 90 名 52%
- ・利用頻度 11 回以上 : 43 名 25.6%

利用頻度 1 回の機械翻訳システムの訪問者は 44 名で、全体の人数比で 25% に相当する。また、利用頻度 2 ~ 3 回は 40 名で全体の 23% である。4 回以上は 90 名で、全体の 52% である。大雑把にいて、約半数は、時々、利用する使用者で、約半数は何らかの自己の業務の中で機械翻訳を必要とする使用者である。

さらに、毎日の業務のなかで、高頻度に利用する使用者は、実稼働日数の 13 に等しいか超える、11 ~ 15 回以上の使用者が該当し、利用頻度 11 回以上に 25.6% 存在している。上記の利用頻度 4 回以上の 52% は固定的に利用する使用者と考えてよさそうである。

使用者の利用内容

次に機械翻訳システムを利用して翻訳した原文および使用者の履歴およびインタビューから、固定的に利用と時々利用する人々について、それぞれの利用形態について調査を行った。

固定的に利用 (52% 90 名)

語学が若干できるが、プロの翻訳家ほど自由に語学を使いこなすことはできず、海外発信など業務の中に翻訳の必要性が生じている利用者層や部門が存在し何らかの翻訳の支援を必要としている。

これらの部門には海外発信が日常業務となっている事

業企画部門や、海外の出先機関との市場対応に苦慮するサービス部門が含まれる。これらの部門において、プロの翻訳家に依頼すると、費用が発生すること、数日のリードタイムを要し、海外との発信で必要となる即時性の要求に対応しきれないため、むしろ機械翻訳を必要とするグループである。

一方で工場、開発部門を中心に機械翻訳の需要が多い。前者は海外生産拠点からの海外発信の必要性から、十分とはいえない語学力をもとに自力で翻訳の必要性に迫られている。機械翻訳による強力な支援を必要としている。後者は海外の技術情報および特許情報の検索に機械翻訳を使用しており、頻繁にインターネットサーフを行っている。つまり日常の業務活動の中で、機械翻訳システムが必然的に入り込むのではなく、ランダムに機械翻訳システムを使用するグループである。

時々利用 (48% 84 名)

機械翻訳システムに興味を持って一度は訪れるが、機械翻訳には興味は持つものの十分に機械翻訳システムを利用してはいえないグループである。

日常業務の中で海外発信を頻繁には行わないが、時々、生産図面の出図などで翻訳が必要となるグループで、これらには工場、開発部門が多い。

以上のことから、逆に翻訳が主業務で翻訳業務として社内地位を築いているグループからの利用はほとんど存在しない。これは現状の翻訳システムが Web 上でのサービス利用を前提として利便性を持たせており、翻訳エディタやメモリ翻訳機能など業務翻訳者向けの機能を提供していないなども要因として考えられる。

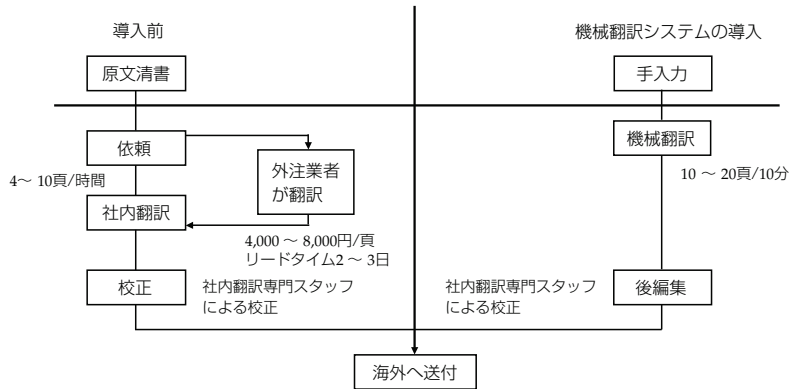


図-4 機械翻訳がもたらした業務プロセスの変革

機械翻訳がもたらす2つの変革と効果

1989年に機械翻訳システムを導入した時点では、高価であったこともあり、使用対象部門は翻訳業務に関係した部門で、そこにはプロの翻訳が存在していた。そのような翻訳家に対して、下翻訳のツールとして使用することを考えた。機械翻訳システムの翻訳性能が十分でないことから、この仮説は正しくなかった。

機械翻訳システムのWeb化によりプロからアマという使用者セグメントの変化が起こり、使用者の増大という機械翻訳システムのブレイクスルーが生じた。

また社内翻訳家に依頼する場合、数時間程度の時間の遅れが生じていたが、図-4に示すように機械翻訳即、後編集という、業務プロセスの変革が生じた。

機械翻訳システムの2つの変革

- ①使用者がプロからアマへ変化
- ②翻訳業務プロセスの変革

以上より、機械翻訳システムの効果は、次の2つにま

とめられる。

- ①英語の翻訳が日常的に発生し、海外の問合せに対して時間を置かずに、応答する必要がある。
 - 即時性の要求に対応
- ②翻訳コストを低減し余った時間を戦略の推敲や仕様書の立案などの創造的な仕事に没頭できる。
 - 知的生産性の向上

現状の機械翻訳システムの品質は決して十分であるとはいえないが、企業内翻訳システムとしては、翻訳を主業務としていないような部門・利用者に対する業務支援ツールとして利用が拡大している。今後、翻訳基本性能が大幅に向上し、第2、第3のブレイクスルーを起こすことができれば、企業の枠を超えて広く世の中全体に受け入れられてゆく可能性を秘めている。

謝辞 本稿の執筆にあたってはご討議いただいたAAMT（アジア太平洋機械翻訳協会）技術動向委員会の委員の方々に深く感謝申し上げます。

(平成15年9月4日受付)

