

翻訳ソフトとマクロによる多言語文書同時作成システムの試作

弁理士 保立 浩一

保立国際特許事務所

1. はじめに

企業活動のグローバル化に伴い、海外での特許取得活動の重要度が増している。各国は自国の言語で特許審査を行うから、当該国の言語で出願する必要があり、特許明細書の翻訳という作業が発生する。このため、多くの弁理士は、作成した特許明細書を英語等に翻訳する作業にも関与している。通常は翻訳業者に依頼して翻訳させるが、自ら翻訳する場合もある。

特許明細書の翻訳にはかなりのコストがかかっており、企業側には削減したいという強い要望がある。このため、機械翻訳（翻訳ソフト）を使用して特許明細書の翻訳を省力化する試みがかなり前から行われている[1]。筆者も、20年以上も前に、当時開発されていた機械翻訳ソフトを使用して特許明細書の英訳を試みたが、「本発明」を *book invention* と翻訳した時点でこれは使えないと思い、早々に断念した記憶がある。

実際、翻訳ソフトについてのユーザーの評価をネットなどで調べてみると、殆どが使えないという評価である[2][3]。短いありふれた文章の英日翻訳であれば翻訳ソフトはかろうじて使えるが、それ以外、特に特許明細書英訳のような専門技術的な文書の翻訳には全く使えない、というのが一般的な評価であろう。使えないというのは、翻訳ソフトで翻訳された英語を手直しして正しい翻訳とする手間と、翻訳ソフトを使わずに最初からマニュアルで翻訳した場合の手間が変わらないか、又は前者の方が大きな手間になってしまうということである。

それでも、短いありふれた文章であれば無料のウェブ翻訳サービスでもかなり高い精度の英訳ができるし、市販の翻訳ソフトでも、ユーザー辞書の登録や翻訳メモリを使いこなすことで、ある程度の翻訳品質が確保できる。

しかし、文章が長くなった際の、特に日英の翻訳では、構文解析に失敗するため、ほぼ100%の確率で誤訳となる。一例として、特許庁が提供する特許明細書のひな型の一文の翻訳結果を図1に示す。

「イメージ入力装置の中で、」は *among image input devices* と訳すべきといった構文以外の問題は別にして、無料ウェブ翻訳では、後半の *so* 以下の部分で構文解析に失敗している。A社製翻訳ソフトの場合、この部分は *** is done...* となっている。これは、「共」が「さ

れている」と翻訳しており、「共」が辞書登録していないためである。「共する」の英訳として *provide* を登録しておく、*it is provided to practical use as an input part of OCR for POS* となり、完全な誤訳とは言えない内容に修正される。しかし、*in the image input device* が *necessary information* を修飾した状態となっており、この点のミスは致命的である。

(原文)

イメージ入力装置の中で、ハンドスキャナは、入力情報の記載された媒体の形状や媒体上の入力情報の位置を問わず、必要な情報のみを入力できる利点があるので、POS用のOCRの入力部として実用に共されている。



(無料ウェブ翻訳)

In the image input device, the hand scanner has an advantage of being able to input only the necessary information regardless of the shape of the medium in which the input information is described and the position of the input information on the medium, so the input of the OCR for POS is shared with practical use as a department.

(A社製翻訳ソフト)

Because the hand scanner has the advantage that it can input only necessary information in the image input device regardless of the position of input information in shape and the medium, ** is done as an input part of OCR for POS by practical use.

図1 現状の機械翻訳の例

翻訳ソフトを使って特許明細書の英訳を行う場合、ユーザー辞書の登録を頻繁に行って翻訳品質を高めることが必須となるが、それでも構文解析のミス、修飾関係の把握ミスがしばしば生じ、その都度、手作業で修正しなければならない。このため、最初からマニュアルで翻訳した方が早かった、という結果になるのである。

多くの翻訳ソフトでは、原文と正しい翻訳文とを対応づけて記憶する翻訳メモリの機能が備わっており、グレードの高い翻訳ソフトでは、一部をワイルドカード(変数)として登録する機能や、任意の一致率の設定を可能にして低い一致率の場合でもヒットするようにする機能も備わっている。このような機能は、同じような文章を翻訳する場合の多い技術マニュアルや取り扱い説明書等の翻訳にはある程度は有効かもしれないが、特許明細書のように以前とは異なる文章の翻訳が主である場合、翻訳メモリの有効性は限定的である。

2. 開発のヒントとコンセプト

このような状況ではあるものの、筆者は、特許明細書のような専門技術的な文書の英訳を翻訳ソフトの利

用により大幅に省力化することを目差し、実用的なシステムの開発を意図した。この理由の一つは、近年における翻訳ソフトの機能向上には目覚ましいものがあり、「本発明」と book invention と翻訳していた時代とは隔世の感があるからである。特許明細書のような技術文書の場合でも、きちんとユーザー辞書を登録し、文章をできるだけ短くしていけば、驚くほど誤訳の少ない結果が得られるようになってきている。

もう一つの理由は、翻訳ソフトの拡張性の一つとして一般化しつつある Office 連携機能である。最近の翻訳ソフトは、Word や Excel といった Office 上のアプリのアドインとして利用できる機能が備わっており、Word 上で入力した文章を Word 上で翻訳したり、あるセルに入力した文章をそのまま Excel 上で翻訳したりすることが可能となっている。Office が使えるということは、マイクロソフト社が Office とともに提供しているマクロ言語である VBA (Visual Basic for Applications) が使えるということである。VBA が使えれば、翻訳ソフトで足りない分を VBA で補うことができ、VBA 程度であれば、筆者のような専門のプログラマーではない者でも、何とか扱える。であれば、実用に耐え得る翻訳も行えるようになるのではないかと考えた次第である。

(1) 文章の各要素のセル化とセル毎の翻訳

短い文章なら精度の高い翻訳ができるのであれば、長い文章を幾つかの要素の分割し、要素毎に翻訳をし、翻訳後に各要素を結合すれば良い。イメージ的には、図2のような感じになる。

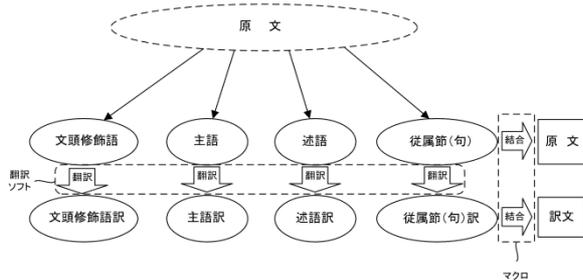


図2 要素毎の翻訳

要素毎の翻訳のメリットは、上述したような修飾関係のミスが基本的に生じないということである。要素内での修飾関係のミスは生じ得るが、要素をまたぐミス、つまり他の要素中の語を修飾する形の翻訳は絶対にされないため、上記のようなミスは生じ得ないということになる。つまり、「イメージ入力装置の中で、」を他の要素から切り離してそれだけで翻訳させ、文全体を修飾する文頭修飾句として後で結合すれば良いの

である。

(2) 各要素の機能のユーザーによる指定

要素毎に分けて翻訳し、翻訳後にマクロで結合するとしても、結合に際しては当該要素の文章中での位置づけの情報が必要になる。つまり、その要素が主語なのか、述語なのか、文頭修飾なのか、それと従属節なのか、といった情報（以下、文内機能情報という。）である。これがないと、マクロは、翻訳後の各要素の結合ができない。

文内機能情報は、翻訳をする者がマクロに対して与える必要がある。このためには、主語ならこれ、述語ならこれ、といったように各機能に対して識別記号を設定し、識別記号をとともにマクロに渡す構成が考えられる。しかし、その分だけ翻訳者の労力が増えてしまい、省力化の意図とは真逆になってしまう。

これを回避するための構成として、各要素を個別の入力欄（テキストボックス）に入力するようにし、入力欄の識別情報（VBA で言えば「オブジェクト名」）を文内機能情報として扱うのである。イメージ的には、図3のような構成となる。

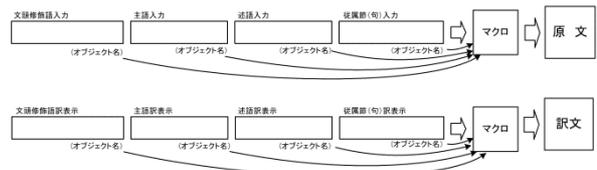


図3 マクロへの文内機能情報の渡し

ただ、この場合も、翻訳者は、各要素を各入力欄に入力していかなければならない。原語の文書の各文章について、各要素を各入力欄に copy & paste していくという作業が発生する。全てをマニュアルで翻訳する場合に比べれば、この部分は微々たる作業であるが、筆者は、ここでもう一段上のソリューションを意図した。

翻訳される原語（文章）も、元々は誰かが作っている。文章の作成者は、自らテキスト入力しているのであり、元々のテキストの入力者が各入力欄に入力するようにすれば、全体として作業量の増加はない。

つまり、日本語での文章の作成の際に最初から要素に分けて入力し、各要素を翻訳ソフトで英訳させた後、日本語、英語それぞれについてマクロで結合すれば良い。この場合、日本語の文章と英語の文章とが同時に出来上がる。つまり、日英文書同時作成である。

3. 構築したマクロ

上記のようなコンセプトの下、マクロを自作し、日英の多言語文書同時作成システムを試作した。

(1) 基本構成

このシステムは、基本的には Excel を利用しており、翻訳ソフトの Excel 連携機能と VBA で記述したマクロを追加している。図 4 に示すように、動作画面は原語（日本語）の入力フォームと訳語（英語）の表示フォームを含んでいる。図 4 の例は、基本文型としての S+V の文型で入力する例となっている。図 4 に示すように、原文の各要素の入力欄と訳文の各要素の表示欄が一对一で表示される画面レイアウトとなっている。



図4 基本構成

ユーザーは、自分が作成しようとしている文章の主語、述語、文頭修飾語、従属節（句）を頭に思い浮かべ、それぞれ入力欄に入力する。そして、翻訳ボタンをクリックすると、翻訳ソフトが起動し、各入力欄の日本語が翻訳ソフトにより翻訳され、その結果が訳語の表示フォームに表示される。

訳語の表示フォームは編集可能となっており、ユーザーは、翻訳ソフトの翻訳結果をマニュアル修正できる。適宜マニュアル修正をした後、結合ボタンをクリックすると、その時点で原語入力フォームで表示されている各要素がマクロにより結合されて原文が生成され、生成された原文が原文表示欄に表示されるとともに原語用のシートに記録される。同時に、訳語表示フォームに表示されている各要素がマクロにより結合されて訳文が生成され、生成された訳文が訳文表示欄に表示されるとともに訳語用のシートに記録される。

各文章の各要素の入力、要素毎の翻訳、各要素の結合による原文・訳文の生成、記録を逐次行い、全ての文章の生成、記録が終わったら、各シートの各行に記録されている文章を結合させることで、日本語と英語の双方で文書が出来上がる。つまり、日本語で文章を入力して文書を完成させると、同時に英語版の文書も出来上がっているということである。

(2) 使用できる文型

基本型の他、複数述語や複数目的語といった他の各種文型でも入力、翻訳が可能となっている。文型の選択は、左側のメニューコラムに設けられた文型選択ボタンで行える。文型選択ボタンをクリックすると、選択可能な文型一覧が表示され、そこから一つを選ぶと、選んだ文型の原語入力フォーム、訳語表示フォームに切り替えられるようになっている。

(3) 要素単位の翻訳メモリの利用

今回試作したシステムで最も効果を発揮したのが、要素毎の翻訳メモリ登録である。翻訳メモリは、本来は一つの文全体として原文と訳文とを対にして登録するものであるが、このマクロでは、各要素の原語と訳語とを対応させて登録するのに使っている。例えば、図 4 の基本型の場合の訳語表示フォームで、文頭修飾語の下の翻訳メモリ登録ボタンをクリックすると、その時点で表示されている文頭修飾語の原語と訳語とが対となって翻訳メモリに登録される。前述した例で言えば、「イメージ入力装置の中で」の対訳として"among image input devices"を登録することになる。

同じような表現が繰り返されると想定される場合、その表現を最初に使用して訳語をマニュアル修正した際に修正後の訳を翻訳メモリ登録しておく、以後はマニュアル修正の必要はほぼなくなる。筆者がこのシステムを使用して実際に日英文書同時作成を試みたところ、要素単位の翻訳メモリ登録が省力化に非常に効果的であることが実感された。

(4) 翻訳の補助ツールとしての利用の考慮

日英同時作成の他、既に出来上がっている文書を英語にする、つまり翻訳業務を行う場合を想定した好適なソリューションも構築した。

このマクロを翻訳業務に使用する場合、原語入力フォームの各入力欄に各要素のテキストを入力する（copy & paste する）という作業がどうしても伴う。これを少しでも省略化するための構成を用意した。

翻訳業務に使う場合でも、ユーザーは文型を予め選択する必要があるが、文型選択の際、文章中に簡単な識別子（区切り記号）を挿入することで原語入力フォームの各入力欄に自動的に貼付がされるようにした。図 5 に、区切り記号を挿入する画面の一例を示す。

ユーザーは、翻訳する文全体をテキストボックスに copy & paste する。そして、文頭修飾語を#b で囲み、主語を#s で囲み、述語を#v で囲み、従属節を#j で囲む。その上で、OK ボタンをクリックすると、マクロにより、各要素が基本型の原語入力フォームの各入力欄に自動的に貼り付けられる。



図5 文区切り入力

各要素のテキストをマニュアルで各入力欄に copy & paste するのとそう変わらない労力にも思えるが、マニュアルの場合、copy と paste でマウスの操作が2回必要である。上記各識別子の挿入の場合、1回で済む。

4. 評価

日本語での文書作成終了と同時に英語版も完成しているというのが理想型であるが、そのために負荷がどの程度増すかが問題となる。この点を確認するため、架空の特許出願依頼についてこのシステムを使って特許明細書の日英同時作成を試みたところ、全体としては20～30%程度の負荷アップであった。

この負荷アップをどうみるかは評価の分かれるところであるが、特許明細書の作成料の例えば30%で英語版の特許明細書が得られるとすれば、現状の英訳料の相場感からすると、格安である。まだまだ改良の余地は多々あるが、筆者個人としては、このシステムのポテンシャルの大きさを実感した次第である。

5. おわりに

試作した日英文書同時作成システムを使用して日英の特許明細書の作成や日→英の翻訳等をしている際、一つ気がついたことがある。それは、常に英語になった状態を想像しているので、日本語での文章を考えている時点で、英訳した際に修正が少なくなる表現を選択していることである。ある内容を表現しようとした場合に日本語での言い回しが頭に幾つか浮かぶが、英語になった時の状況を考慮して瞬時に幾つかを除外し、英語にした場合に最適な言い回しを選んで日本語入力するのである。これは、ある意味、海外生活が長い人

が日本に帰ってきて日本語を話すのに似た状況である。

例えば、イチロー選手のインタビューを聞くと、日本語で話しているが、その構文は完全に英語だなど感じる事が多い。頭の中は完全に英語で（つまり、英語の質問に英語で答えている状況で）、アウトプットの最後の段階で各要素を日本語に置き換えている。例えば、何かのインタビューで彼は「これは僕にとって難しい決断でした。」と答えていたが、これは、頭の中は”This was a difficult decision for me.”となっていて、それが口から出る寸前に日本語に置き換えている。これが、普通の日本人であれば、「色々と悩んで、やっと決断しました。」というような言い回しになるであろう。

日英文書同時作成システムを使用している時も、同じように感じる時がある。頭の中は完全に英語モードになっていて、キーボードを叩く段階で日本語にしている。つまり、頭の中で英語を日本語に逆翻訳し、それからキーボードを叩いている。そのようにしないと、英語の修正が多くなってしまい、日英同時作成のために労力が過大となってしまうからである。例えば、イチロー選手の例で言えば、「これは僕にとって難しい決断でした。」を英訳するのと、「色々と悩んで、やっと決断しました。」を英訳するのと、どちらが誤訳が少なくなるかは、自明であろう。

言語処理の分野にはニューラル機械翻訳というのがあるそうだが、日本人が日本語の文章を作る際の脳プロセスと欧米人が英語の文章を作る際の脳プロセスとは違う筈で、その違いを乗り越える必要があるように、門外漢ながら感じるところである。もっとも、日本人のだれもが「色々と悩んで、何とか決断しました。」と言わずに「これは僕にとって難しい決断でした。」と言うようになる、つまり日本人の脳プロセス自体が英語化してしまえば、乗り越えるべきハードルも低いものになる。ただ、そのような状況を生み出すには、英語を第二公用語にするぐらいのことをしなければいけないようにも思うのである。

参考文献

- [1]熊野明. 特許明細書の日英機械翻訳における課題, Japio 2008 YEAR BOOK pp.88-92
- [2]<http://review.kakaku.com/review/newreview/CategoryCD=0365/>
- [3] <https://www.translator.jp/soft/>