

翻訳ソフトとマクロによる特許明細書の日英同時作成



会員 保立 浩一

要 約

翻訳ソフトとマクロを利用して特許明細書を日本語及び英語で同時作成するための手法について研究した。翻訳ソフトは Office 連携機能を有するものであり、マクロは VBA(Visual Basic for Applications)で自作した。原文の各要素をセル化し、セル毎に翻訳ソフトで翻訳させた後に結合する手法を用いた。各要素のパターンは、ユーザーが選択する文型に応じたものであり、文型に応じたパターンで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した。日本語での特許明細書作成時の負荷は 20~30%程度増えるものの、高い翻訳精度の英語版の特許明細書を同時作成できることが確認できた。特に、要素単位で原語と訳語とを翻訳メモリ登録をして活用する構成は、省力化に大きな威力を発揮することが実感された。

目次

1. はじめに
2. 開発のヒントとコンセプト
 - (1) 文章の各要素のセル化とセル毎の翻訳
 - (2) 各要素の機能のユーザーによる指定
3. マクロの概要
 - (1) 基本構成
 - (2) 使用できる文型
 - (3) 要素単位の翻訳メモリの利用
 - (4) 翻訳の補助ツールとしての利用の考慮
4. 評価
5. おわりに

1. はじめに

経済のグローバル化に伴い、企業の知財活動も海外の占める割合が大きくなってきている。海外での知財活動では、紛争対応の重要性が増しているが、その前提として、防衛手段となる権利取得も依然として大きな要素となっており、多大なコストを占めている。

海外での権利取得に要するコストのうち翻訳料は無視できない部分であり、特に特許明細書の翻訳には多大な費用がかかっている。翻訳業者間での競争もあり、以前に比べるとかなり安価となってきているが、それでも品質の良い翻訳サービスを利用すると、安くはない費用が請求される。

一方、世の中には機械翻訳(コンピュータ翻訳)を行う翻訳ソフトが出回っている。10万円を越える高

価なものもあるが、安価なものでは数千円で手に入るものもある。このような翻訳ソフトを使用して特許明細書の翻訳をすることで、翻訳費用を大幅に低減できるのではないかと考えられ、実際に試してみた実務家の方も多であろう。

筆者自身も、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。特に日英の翻訳については、非常に厳しい評価となっている。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、それ以外では、特に特許明細書英訳のような専門技術的な文章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、無料のウェブ翻訳サービスでも、短いありふれた文章であれば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を英訳してみよう。無料ウェブ翻訳、A 社製翻訳ソフトのそれぞれについて、英訳の結果を図 1 に示す。尚、A 社製翻訳ソフトでは、ユーザー辞書は特に追加登録していない状態である。

(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 It 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

「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、構文上の問題として、無料ウェブ翻訳では、後半の so 以下の部分において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。

A 社製翻訳ソフトのような翻訳ソフトを使って特許明細書の英訳を行う場合、上記のような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好適かもしれないが、特許明細書には向かない。特許明細書の場合、過去の案件と内容が違うから特許を出すのであり、必ず内容は異なっている。したがって、特許明細書の英訳については、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、翻訳ソフトの利用による特許明細書英訳の大幅省力化を目差し、効果的な手法の開発を意図した。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、下の図 2 のような感じになる。

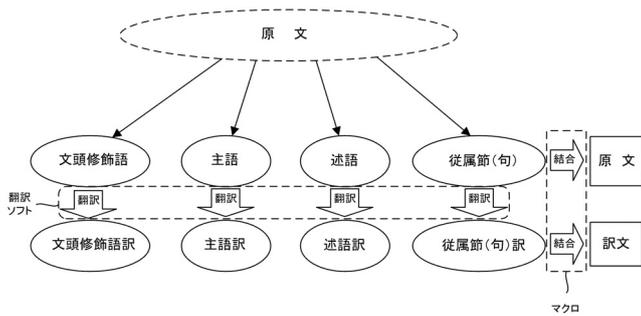


図2 要素毎の翻訳

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ので、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 (=日英文書同時作成)

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、以下の図3のような構成となる。

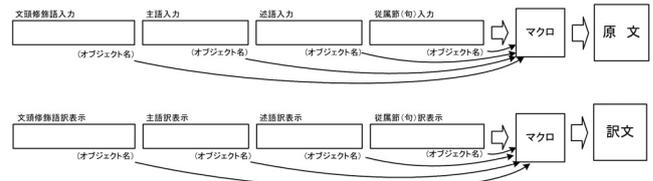


図3 マクロへの文内機能情報の渡し

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。ここでの文書は特許明細書であるから、明細書作成者が各入力欄にテキスト入力をすれば良いのである。

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。

内外出願を受任する弁理士の場合、日本出願の受任時に特許明細書を作成し、海外出願時には英訳を翻訳業者に依頼する。そして、納品された特許明細書の英訳をチェックし、必要な修正を指示する。この場合、最初から海外出願が予定されている案件では、英訳をイメージしながら日本語の特許明細書を作成する場合も多いであろうし、誤訳がないように構文や日本語を選びながら作成する場合も多いであろう。上述した日英文書同時作成は、日本語での特許明細書作成と英訳された特許明細書のチェックとを同時に(逐次に)行っているような状況であり、それほど違和感はないように思われる。

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、特許明細書の日英同時作成を試みた。以下、概要を紹介する。

図4は、マクロの動作画面の一例を示す概略図である。



図4 基本構成

(1) 基本構成

この手法は、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。これは、原文の各要素に対する翻訳ソフトによる翻訳状況を把握し易くして訳の修正等が容易にできるようにするためである。

尚、この手法では Excel を使っているので、一つの文がシートの一つの行に記録される。原文用のシートと訳語用のシートとが用意され、各要素に対応してセルが割り当てられている。

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、他の各種の文型でも入力、翻訳が可能となっている。以下、作成した文型の一覧である。

- 文型1：欄一つのみ
- 文型2：文頭修飾語 + 本文
- 文型3：基本型
- 文型4：複数述語
- 文型5：複数目的語
- 文型6：主語関係詞
- 文型7：目的語関係詞
- 文型8：単純複文
- 文型9：仮主語文
- 文型10：there 文

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。図5に、他の文型の一例として複数述語の場合を示す。

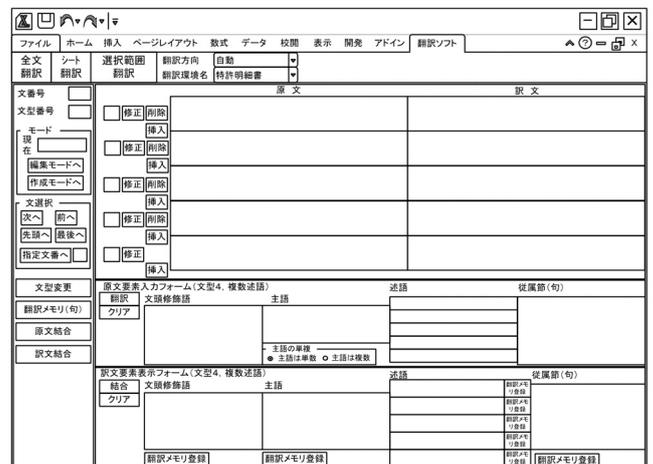


図5 複数述語

(3) 要素単位の翻訳メモリの利用

今回試した手法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。

要素毎の翻訳メモリ登録では、要素は複数の語句から成る場合、その一群の語句の対訳を登録することになる。例えば、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。

特許明細書の場合、内容的には以前のものとは異なるとはいっても、一つの特許明細書の中では、同じような表現が繰り返し使われる。特許明細書では、同じ内容は同じ表現で説明するというルールがあるため、必然的である。したがって、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。筆者がこの手法を使用して実際に特許明細書の日英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ので、以下に紹介する。

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6に、区切り記号を挿入する画面の一例を示す。図6では、基本型(S + V)で翻訳をする場合が示されている。



図6 文区切り入力

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。

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回必要であるが、それに比べると、上記各識別子の挿入の方が断然に楽である。この点も、実際に使用してみた際に実感された。

4. 評価

日本語での特許明細書の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手法を使って特許明細書の日英同時作成を試みた。細かな所要時間の測定はしていなかったが、全体としては、日本語のみで特許明細書を作成する場合に比べて20~30%程度の負荷アップといったところであった。

この負荷アップをどうみるか。例えば特許明細書の作成手数料が30万円であったとした場合、20%の負荷アップで英語版の特許明細書ができたとすれば、プラス6万円で英語版の特許明細書も提供できることになる。30万円の特許明細書について翻訳業者に英訳を依頼した場合、単語数によって異なるから一概に言えないが、少なくとも倍以上の翻訳料は請求されるであろう。つまり、英訳のコストは半分以下となり、しかも英語版が日本語版と同時に提供されるということ

になる。まだまだ改良の余地は多々あるが、筆者個人としては、この手法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

そう遠くない将来、特許明細書のような専門技術的な文書も、AI（人工知能）によって正確に翻訳される日が来るであろう。本稿で紹介した手法も、仮に実用化されたとしても、それまでの過渡的なものとなることは間違いない。

ただ、「イメージ入力装置の中で、」を、“in the image input device”ではなく“among image input devices”と訳すためには、前後の文脈のみならず文書全体を把握した上で「イメージ入力装置の中で、」の意味するところを判断する必要がある。さらに、明示はされていないが、「ハンドスキャナ」が「イメージ入力装置」の下位概念であるという前提としての技術知識も必要である。このような諸々の事項を了解した上でコンピュータが「イメージ入力装置の中で、」を“among image input devices”と訳すためには、まだまだ超えなければならない高いハードルが数多くあるように思う。

話は変わるが、本稿で紹介した手法は、特許審査においても使用できるように思われる。例えば、現在、日本国特許庁を受理官庁として英語でPCT出願をした場合、国際調査機関（ISA）としての日本国特許庁は、英語で国際調査報告（ISR）や国際調査見解書（WO/ISA）を作成する。したがって、日本国特許庁

は一部で既に英語審査をしているといえる。

通常の国内出願で完全な英語出願を認めるかどうか（翻訳文提出を不要にするかどうか）は別にして、審査態勢としては、日本は、英語出願が可能な態勢が取られている、ないしは取られつつあるとみて良いだろう。仮に、通常の国内出願で外国人の出願人用に英語での審査結果を通知する制度を設けた場合を想定してみると、権利の対世的効力や権利解釈における出願経過参酌等を考慮し、行政の記録としては日本語で保存されていることが望まれる。つまり、日本語と英語との双方で審査結果の文書を作ることが必要になるだろう。このような場合、本稿で提案された手法が使用できるのではないかと、手前味噌ではあるが、考える次第である。

尚、英語審査が本格化すると、いよいよ日本国特許庁のハブ特許庁化、特許審査という行政サービスの海外輸出の時代が到来するのであろう。世界一の審査能力（スピード、質）に加え英語審査という武器を備えた時、世界のイノベーションを日本国特許庁が特許審査で支えるという状況が見えてくる。イノベーションに関する情報が集積するという状況は、国内企業によるイノベーションをさらに促進する礎ともなろう。

(注)

Visual Basic, Excel は、米国 Microsoft Corporation の米国その他の国における登録商標又は商標である。本文中、® の表記は省略した。

(原稿受領 2017. 8. 9)