

翻訳できる日本語

一般財団法人日本特許情報機構 特許情報研究所顧問 横井 俊夫

PROFILE

1966年に電気試験所（現在：産業技術総合研究所）。1982年より第五世代コンピュータプロジェクトの推進に従事。1987年より電子化辞書プロジェクトの推進、運営に従事。1995年よりフィリピン国にてODAプロジェクトの推進、指導に従事。1997年より東京工科大学、2008年より現職。



産業日本語は多様な側面を持つ。その産業日本語を翻訳できる日本語という側面から解説する。特許情報の観点からは、特許文章の日本語から英語への参照用翻訳を取り上げる。しかし、説明としては、一般の産業技術文章や英語以外の外国語への翻訳も念頭に置いたものである。翻訳できる日本語は、産業日本語の効用を実感させ、産業日本語に関する技術開発を具体化するための糸口となる。

1 明晰であるとは、翻訳できること

産業日本語は、人にもコンピュータにも明晰な日本語である。「明晰な」は、「効率よく正確に理解できる、書くのも容易な」と言い換えることができる。しかし、「理解」というのは、頭脳の中での現象である。そのような現象について「効率がよい」とか「正確である」とかを議論できるようになるのは、遥かに先のことであろう。

人は、与えられた産業技術文章に様々な知的行為を施すことによって、新たな価値を持つ新たな文章を作り出す。この与えられた文章と作り出された文章は、客観的に観察し、分析することができる。このような文章対が十分な量用意されれば、場合によっては、知的行為そのものも部分的にしる明らかにすることができる、あるいは、できるかもしれない。

人が文章に施す知的行為の典型として、「文章を翻訳する」「文章を検索する」「文章を要約する」を挙げることができる。「文章を翻訳する」とは、その文章から

原言語を母語とする読み手が理解する内容とできるだけ近い内容を訳言語を母語とする読み手が理解できるように原文を訳文に言い換えるという知的行為である。「文章を検索する」とは、文章集合から検索要求文と一致する内容を含む文章を探し出すという知的行為である。検索要求文と文章が一致するか否かの判断は、検索要求文に言い換えうる文を文章が含んでいるか否かの判断である。「文章を要約する」とは、その文章の内容を縮約し、縮約内容を表現する簡略文章に言い換えるという知的行為である。

人が行う文章翻訳、文章検索、文章要約に対応して、コンピュータとしても、実用性のレベルは様々ではあるが、機械翻訳システム、情報（文書）検索システム、文書（文章）要約システムが開発され利用されるようになっていく。そこで、人にもコンピュータにも明晰である日本語とは、人が効率よく正確に翻訳、検索、要約できる文章を容易に書ける日本語、そして、コンピュータの機械翻訳システム、情報検索システム、文書要約システムが効率よく正確に処理できる文章を容易に書ける日本語であると定め直すことができる。そして、ここでは、とりあえず、人が効率よく正確に翻訳できる日本語、機械翻訳システムが効率よく正確に処理できる日本語を取り上げる。

2 翻訳するとは、翻訳原稿を作成すること

人が文章を翻訳するとは、その文章から原言語を母語

とする読み手が理解した内容とできるだけ近い内容を訳言語を母語とする読み手が理解できる訳文に言い換えることである。しかし、ここにも、「内容を理解する」というくだりが含まれている。このままでは、やはりエンジニアリングの対象とするのは難しい。

特許文章の参照用翻訳では、直訳すること、あるいは、原文に忠実に翻訳することが求められる[1]。特許文章がすでに翻訳できる日本語で書かれていれば、「内容を理解する」を含まない、あるいは、含まないといっているほどの翻訳プロセス、すなわち、文字どおりの直訳ですませることができる。しかし、現状の多くの特許文章は、はなはだ非明瞭な日本語で書かれているため、翻訳者は、非明瞭な文章を理解する、時には、謎解きをするような理解作業を強いられることになる。

そこで、翻訳原稿という仲介物を設けることにする。すると、翻訳は、原文から翻訳原稿を作成し、この翻訳原稿を直訳することによって訳文を作成するというプロセスとなる。原文の内容を理解するというステップは、原文から翻訳原稿を作成するというステップに置き換えられる。原文の内容を理解するという翻訳者の知的行為は、行為の入力となる原文と行為の出力となる翻訳原稿の対によって観測し分析することが可能となる。

そして、翻訳原稿からの翻訳は、機械翻訳に任せる、あるいは、機械翻訳を利用して効率化することができる。翻訳者の翻訳作業の大半は、翻訳原稿の作成作業に置き換わる。このようにすることによって、翻訳プロセス全体がエンジニアリングの対象となる。

3 特許翻訳のプロセス

翻訳原稿を介することによって、特許ライティング（特許文書ライティングと特許文章ライティングを含む）において目指すべき特許翻訳はどのようなものだろうか。図1に特許ライティングのプロセスを示す。この図は、特許文書・特許文章のライティングに焦点を絞ったものであり、発明の着想・抽出のフェーズや

特許調査のフェーズ等は省略されている。また、将来的には、産業日本語プラットフォームや特許情報クラウドという新しい環境の下で特許翻訳を含む特許ライティングのプロセスは大きく構成し直されるであろう。ここでは、翻訳原稿ともうひとつの仕組み（とりあえず、「翻訳情報ベース」と呼ぶことにする）とによって、特許翻訳業務が如何に改善されるのかである。

翻訳業務のプロセスは、翻訳原稿と翻訳情報ベースを介することによって以下ようになる。

- (1) 依頼者から受注し、依頼原稿を受け取る。
依頼原稿は、特許ライティングにおける完成原稿とほぼ同じである。
- (2) 依頼原稿から翻訳原稿を作成する。
翻訳者からの依頼者への問い合わせを含む。
- (3) 翻訳原稿を訳する（機械翻訳を活用する）。
翻訳者は、翻訳原稿を直訳に近い形で翻訳する。
- (4) 翻訳結果を整える。
出願国の特許文書・特許文章特性に合うように整える。
- (5) 翻訳情報ベースを更新する。
翻訳プロセスの各段階で逐次に更新される場合もある。
- (6) 依頼者に訳文、翻訳原稿、依頼者別翻訳情報ベース（その更新データ）を納品する。
依頼者は、訳文のチェックを翻訳原稿のチェックで代替できる。なお、図1で「+α」と印したものに、翻訳原稿や更新データが含まれる。

ここで、翻訳原稿と翻訳情報ベースの役割を確認しておく。翻訳原稿は、翻訳できる日本語で書かれ、依頼原稿が明瞭な日本語で書かれている場合は、手順だった作業によって作成される。そして、以下のような役割を担う。

- ① 翻訳のプロセスを外在化し分析可能にし、翻訳の知識を蓄積可能・再利用可能にする。
- ② 翻訳に機械翻訳を利用できるようにする。

- ③ 翻訳結果のチェックを翻訳原稿のチェックによって代替できるようにする。

次に、翻訳情報ベースの役割である。翻訳情報ベースは、翻訳に係る言い換え規則や複合語や対訳コーパスなどのデータ・情報・知識を統合的に管理する情報ベースである。共通翻訳情報ベースと依頼者別翻訳情報ベースに分かれる。共通翻訳情報ベースは、翻訳事業者等の翻訳専門機関が利用・管理する翻訳情報ベースである。依頼者別翻訳情報ベースは、依頼者機関ごとに設けられ、依頼者の機密事項や著作権に係る翻訳情報をまとめた翻訳情報ベースである。依頼者別翻訳情報ベースの管理は、依頼者自身が行う場合と翻訳専門機関に委託する場合がある。

4 特許ライティングマニュアル

翻訳できる日本語の仕様は、特許ライティングマニュアル（図1）にライティング規則として具体化される。ライティング規則は、原則として言い換え規則の形式にまとめられる。特許ライティングマニュアルは、大きく、和文原稿作成用と翻訳原稿作成用に分かれる。和文原稿作成用は、その分野の一般の技術者や研究者が、自分達にとって十分に明晰な和文原稿を作成できるように

するためのものである。明晰な和文原稿に基づく依頼原稿から、翻訳原稿作成用に従う手順だったプロセスによって、翻訳原稿が作成される。翻訳原稿作成用は、対象とする外国語共通となる部分と外国語個別となる部分に分かれる。したがって、翻訳できる日本語は、日本語として明晰な日本語と直訳できる日本語とから成り立つことになる。そして、直訳できる日本語は、英語に直訳できる日本語、中国語に直訳できる日本語等々に分かれることになる。

特許ライティングマニュアルは、現在、Japioの特許産業日本語委員会のもとに設けられた特許ライティング・タスクフォース（PWTF）が日英翻訳を対象に策定を進めている。本年10月末には、第0版をまとめ、来年3月末には、第1版をまとめ広く公開する予定である。

5 文の容認性という明晰性

翻訳できる日本語の策定は、特許管理、特許翻訳、機械翻訳、言語処理、特許情報等の多角的な観点から進められている。そこでの議論の一端を紹介しておこう。文の容認性に関する議論である。これは、日本語として明晰な日本語に係る議論である。

文レベルの明晰性の指標として、文の構文構造の容認

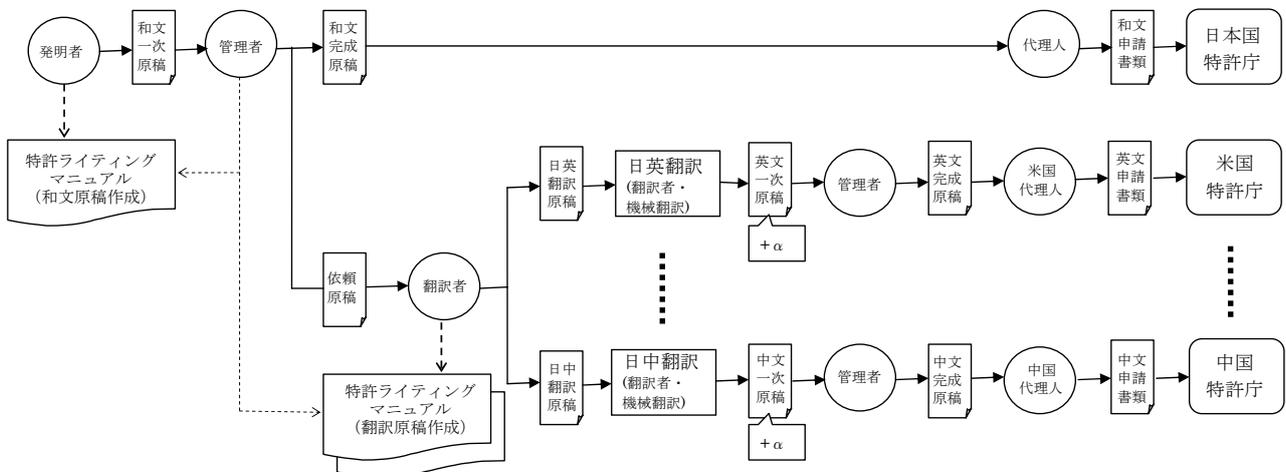


図1 特許ライティングのプロセス

性がある。言語学では、文の文法性 (grammaticality) と文の容認性 (acceptability) が議論される [2]。文がその言語の文法において適格であるかどうかを判断する基準となるのが文法性であり、文がその言語の読み手・書き手にとって解釈可能 (解釈容易) であるかどうかを判断する基準となるのが容認性である。

容認性については、キンボル (Kimball) の原則があり、原則のひとつとして、二文の原則 (Two Sentences Principle) がある。文は、通常、主節と複数の従属節によって構成され、複数文の入れ子構造となる。この入れ子構造に関して、同時処理する文の数が3以上になると著しく容認性が低下することから、同時処理する文の数を上限2に抑えよという原則である。これは、心理学で指摘される事実、すなわち、人間の作業メモリ (短期記憶) の容量が限られているという事実由来する原則である。したがって、この原則は、すべての言語に適用できる。

ただし、日本語と英語では、文の基本構造が異なるため、この原則に反する和文を翻訳した場合に翻訳英文がこの原則に反することになるわけではなく、逆に、この原則に反する英文を翻訳した場合に翻訳和文がこの原則に反するようになるわけではない。例えば、以下の英文は、文としては、決して長い文ではないが、3つの文が中央埋め込み (center-embedded) の入れ子構造となるため、非常に容認性が悪くなる。二文の原則に反する英文である。

- a. The student the professor the scientist collaborated with had advised copied the article.'

以下のように関係代名詞を明示したとしても事情は変わらない。

- b. The student who the professor who the scientist collaborated with had advised copied the article.'

そして、この英文を直訳すると以下の和文となる。こ

の和文は、二文の原則に沿う、この原則に限れば明晰な文である。

- c. 「科学者が一緒に研究した教授が助言した学生がその論文をコピーした。」

逆に、この和文に沿った英文を作ると、二文の原則に沿う英文となる。

- d. 'The scientist collaborated with the professor who had advised the student who copied the article.'

一方、以下の和文は、文としては短い文であるが、二文の原則に反するため非明晰である。

- e. 「科学者が教授が学生がコピーした論文を渡した研究者を褒めた。」

この和文を直訳すると以下の英文となる。この英文は、二文の原則に沿う明晰な文である。

- f. 'The scientist applauded the researcher whom the professor gave the article which the student copied.'

キンボルの二文の原則は、従属節ばかりではなく複雑な並立表現や複雑な修飾句を含めると二構造の原則 (Two Structures Principle) へと拡張することができる。すなわち、同時に処理可能な構造体の上限は、2であるという原則である。そして、特許文章には、この二構造の原則に反する中央埋め込み文が頻繁に用いられる。特許文章を非明晰にしている大きな要因である。

特許文章では、請求項を一文で書く慣習がある。しかも、請求項の文章が、ほぼそのままの形で特許文書の随所に用いられる。ただし、請求項の文章形式は、慣習化されており、その骨格構造はほぼ定型化されている。そのため、特許の専門家にとっては、容認性の障害とはならないという意見もある。

以下は、実際の要約書の文章例である。

【要約】

「ラベル剥離ユニットは、用紙搬送装置の筐体に取り付けられる取付体と、取付体に対して係合しながら移動可能であり、ばねで付勢されることによって取付体に対して閉じた状態又は開いた状態において停止する可動体と、可動体に取り付けられ、用紙搬送装置によって搬送されるラベル連続体の台紙が当接されて台紙を湾曲させることにより、台紙からラベルの一部分を剥離するラベル剥離体と、可動体に回転自在に取り付けられ、台紙を介して用紙搬送装置のプラテンローラに接して回転することにより、台紙を搬送して外部に排出する台紙搬送ローラとを具備する。」(257文字)

この文は、以下に示す基本構造だけで、2構造の中央埋め込み文である。そして、並立表現要素のいくつかは、3構造にもなる埋め込み構造である。

「ラベル剥離ユニットは、～取付体と、～可動体と、～ラベル剥離体と、～台紙搬送ローラとを具備する。」

一方、以下は、要約原文の英訳である。英文の埋め込み構造は、並立表現を除いて中央埋め込みではない。しかも、並立表現は、セミicolonという特別の区切り記号によって並立要素が明示化されている。さらに、英語では、修飾句や修飾節が被修飾名詞の後ろに置かれる。この英文は、二構造の原則に関する限りは十分に明晰である。

ABSTRACT OF THE DISCLOSURE

'A label peeling unit having
an installation body installed on a housing
of a paper sheet conveyance device;
a movable body movable while engaging
with the installation body and stopping in a
closed state or an open state in which the
movable body is closed or open relative to
the installation body;
a label peeling body installed on the
movable body and allowing backing paper of

a continuous label body, which is conveyed
by the paper sheet conveyance device, to
be in contact with the label peeling body to
bend the backing paper, peeling a part of a
label off the backing paper;

and a backing paper conveyance roller
rotatably installed on the movable body
and rotating while being in contact with a
platen roller of the paper sheet conveyance
device with the backing paper in between
to thereby convey the backing paper and
discharge it to the outside.'

先の【要約】の文章は、独立請求項の文章をほぼそのまま転記したものである。そして、その英訳であるABSTRACTの文章は、同じ独立請求項の英訳文をほぼそのまま転記したものである。ところが、構造の埋め込みに関して、英文の方がはるかに明晰である。

請求項に関して、米国特許は、細かな表現形式を定めている。その表現形式は、厳密に言えば文の形式から逸脱しているが、広義には一文表記の形式である。文とすれば、非常な長文である。長文という点に関しては、日本特許の請求項と変わらない。ところが、この長文が、二構造の原則に反しない、あるいは、ほとんど反しない構造となるように表現形式が定められている。先の例に沿えば、aやbのような表現形式ではなく、dのような表現形式であるということである。さらに、修飾-被修飾関係の曖昧さを避けるために関係代名詞による関係節の使用を控え、準動詞句を使用することが推奨されている[1]。例えば、dのような文を以下のgのような形式にすることが推奨されている。

g. 'The scientist collaborating with the
professor advising the student copying
the article.'

日本特許の請求項に関する表現形式は、慣行として長文の一文形式をとる。この一文形式のいくつかの形式は、米国特許の請求項に関する表現形式の翻訳から得られた

ものであろう。ところが、日本特許の表現形式は、先の【要約】の文章例に見るように、二構造の原則に大いに反するものになってしまっている。先の例に沿えば、fである英文をeとなる和文としたようなものである。

英文の表現形式を模するあまり、英文の明晰さを模するのではなく、至って非明晰な和文の表現形式を作りだしてしまったのである。このような和文の表現形式に慣れた特許の専門家には、さほどの問題ではないと言う説もあるが、実際の特許文章を仔細に分析してみると基本的な日本語文章としての不適切さがかなり目立つ。特許の専門家にとっても、このような複雑な長文を扱うのは、かなりの負担になっているのではなかろうか。請求項の文章までとなると抵抗があるかもしれないが、明細書や要約書に使用される請求項文章の転記文は、日本語として明晰な文に言い換えても良いのではなかろうか。例えば、先の【要約】の文は、以下のような連文表現に言い換えることによって、格段に明晰となる。この言い換えは、手順だった機械的な処理で行うことができる。

【要約】

「ラベル剥離ユニットは、取付体と可動体とラベル剥離体と台紙搬送ローラとを具備する。

取付体は、用紙搬送装置の筐体に取り付けられる。可動体は、取付体に対して係合しながら移動可能であり、ばねで付勢されることによって取付体に対して閉じた状態又は開いた状態において停止することができる。

ラベル剥離体は、可動体に取り付けられ、ラベル連続体の台紙が当接されて台紙を湾曲させることにより、台紙からラベルの一部分を剥離する。なお、ラベル連続体は、用紙搬送装置によって搬送される。台紙搬送ローラは、可動体に回転自在に取り付けられ、台紙を介して用紙搬送装置のプラテンローラに接して回転することにより、台紙を搬送して外部に排出する。」

(303文字)

[参考文献]

- [1] 倉増 一：特許翻訳の基礎と応用 - 高品質の英文明細書にするために -、講談社（2006）
- [2] 磯部美和：ことばの理解のメカニズムをさぐる、はじめて学ぶ言語学 - ことばの世界をさぐる 17章 - （大津由紀雄編著、ミネルヴァ書房）、第9章、pp.169-190（2009）