

実務日本語とUTXから 特許明細書の改善を考える

秋桜舎代表 **山本 ゆうじ**

PROFILE

筑波大学を経てシカゴ大学修士号。実務翻訳業務、大規模翻訳・文書管理／作成の講習やコンサルを行う。近著に「IT時代の実務日本語スタイルブック——書きやすく、読みやすい電子文書の作文技法」。

✉ <http://transpc.cosmoshouse.com/> (連絡用フォームから)

1 実務日本語の「百半ルール」

本記事では、特許明細書での日本語作文を改善するための具体的な方法として実務日本語の「百半ルール」、用語集形式 UTX、JTF スタイル ガイドについての概要を紹介する。

現状の特許明細書は、慣習的に、書き手の側に「簡潔に分かりやすく書こう」という意識が見られず、文章としてさまざまな問題を抱えている。

特許明細書の主な問題点は、以下の3点である。

- 長い（一文が長く、文書全体も長い）
- あいまい
- 難しい

これらは、特許明細書に限らず、すべての日本語の実務文章全般に共通する問題でもある。これらの問題点は、相互に関連している。文が長いことは、文章をよりあいまいにし、難しくしている。

また、上記の3つの問題点は、より具体的な、さまざまな問題点を生む。まず、表記のあいまいさにより、読み手の側に誤解が生まれる。さらに、文書をデータベース化した場合に、検索がしづらくなる。これは結果的に、日本語の特許での国際的な競争力を低下させる。また、人力翻訳と機械翻訳での翻訳の費用と時間を増加させ、精度を減らすことになる。

このような問題の原因はなんだろうか。以下の点が考えられる。

- 特許明細書の関係者で、難解な文章がもたらす問題について、切迫した危機意識が薄い

- 文書作成者が、構造的文書を作成する訓練を受けていない
- 文書作成者が、自分の文書の問題点を認識できていない
- 問題点を解決するための、具体的で実行可能な方法を実施していない
- 問題があった場合、それを指摘して修正する制度がない
- 難解な文章に対する、無意識のあこがれ、心理的・知的優越感がある
- 学校での国語教育が稚拙かつ旧態依然である
- 「簡潔に分かりやすく書く」ことに対して、直接的な動機付けがない
- 学術的な研究と現場との間に接点がなく、実践として結びつかない
- 具体的な方法がないと、「分かりやすい書き方」の実践のしようがない

特許文章での最重要かつ本質的な問題は、技術的な問題ではない。つまり、どのような特許文章作成支援ソフトがあっても、それだけでは解決しない。特許文章が、不必要に長く難解なのは、慣習に根ざす心理的抵抗と、「どのように明解に書くか」という具体策が取られないことによるものである。特に長い文を避けるには、文を完結に短く書くことの利点と、文を長く書くことによる多大な労力と費用の増加を明確化して、関係者の共通理解とする必要がある。

では、具体的にはどうしたら「長い・あいまい・難しい」文章の問題を解決できるのか。まず、特許明細書を改善するための、実行可能な方法として実務日本語の「百

半ルール」(山本ゆうじ著『IT時代の実務日本語スタイルブック——書きやすく、読みやすい電子文書の作文技法』p. 90)について紹介する。実務日本語の百半ルールのみで上記の問題のすべてを解決できるわけではないが、いくつかは改善できる。

「百半ルール」とは、私が提唱する「実務日本語」の基本ルールの一つであり、「1文が100字を超えたら半分にする」というシンプルなルールである。単純なようだが、以下の利点がある。

- 理論のみでなく現実に実践可能である
- 「長い・あいまい・難しい」の問題点を同時に解決できる(作文、人力翻訳、機械翻訳のどの場合でも)
- チェックに文法知識は一切不要である
- ツールによる自動チェックができる
- 長い専門用語を含む文でも、例外をもうけることなく、無理なく、徹底して施行できる

「長い文を短くすればよい」ということ自体は、以前から言われており、それ自体は目新しくはない。たとえば、オンダ国際特許事務所の恩田博宣氏は、明細書で「3行を超えない書き方」を提案している。だが、長い文とはどこからが長い文なのか、具体的に何字以下であればいいのか、どのようにすれば確実に実施できるかについては不透明な点があった。

特許版・産業日本語委員会 報告書では、「推奨連文形式」(『平成23年度 特許版・産業日本語委員会 報告書』PDF上でのp.130)が提案されている。百半ルールは、これを具体的に実践する1つの手法である。

本来であれば、特許明細書から100字以上の実例を示すことが適切である。これは著作権に関する適切な「引用」の範囲内であり、支障はないはずだが、特許版・産業日本語委員会からの要請により、今回は特許明細書の例を避け、長い文例の宝庫であるe-Taxサイトから例を採る(<<http://www.e-tax.nta.go.jp/e-taxsoftweb/e-taxsoftweb1.htm#Link6>>)。このサイトには以下のような長い文がある。

「e-Taxでは、申告等データに電子署名と電子証明書を添付することによって、申告等データが利用者本人の作成したものであることや改ざんのないことを確認しま

すが、あらかじめ電子証明書をe-Taxに登録しておくことにより、その確認のための処理を短時間で言い、送信者への受信通知を速やかに送信することとしています。」

この文は152文字あるが、1文で表現しなければならない必然性はまったくない。ここでは2文に分割することで、理解しやすく、翻訳しやすくできる。

この文は、以下のように分割できる。

「e-Taxでは、申告等データに電子署名と電子証明書を添付することによって、申告等データが利用者本人の作成したものであることや改ざんのないことを確認します。(←77文字) あらかじめ電子証明書をe-Taxに登録しておくことにより、その確認のための処理を短時間で言い、送信者への受信通知を速やかに送信することとしています。(←74文字)」

この文には他にも問題がいくつもあるが、ここでは単に分割しただけである。この作業自体はシンプルであり、負担にはならない。作業を理解するのに、文法用語は必要もない。実際に数万語にもなる多量の文書を読みやすく修正する場合、なるべく単純な作業でなければ実施は困難である。その点でも、百半ルールは現実的といえる。また分割後の文でも、「ことによって」「することとして」などと回りくどい表現が残っているが、読みやすくするには、より直接的な表現に言い換える必要がある。

百半ルールの具体的な方法として、テキストエディターでは正規表現、Microsoft Wordでは以下のワイルドカードを使う(全角と半角に注意)。

。[!。!?^13]{100,}。

このワイルドカードは簡易的な方法なので、よりよいものに改善できる余地がある。この記事自体でも、割り注を含む2文以外は、100字を超える文は1文もないことを確認している。

百半ルールのポイントは、「字数を削って100字以下にするのではなく、複数に分割する」ということである。元が100字超の文を2つに分割した後の字数は、半分程度になる。100字を超える文は、複文(1文中に、主語と述語の関係が複数回、入れ子構造になっている文)である可能性が高い。そのため、字数を削るより



も、文を分割することにより、わずかな編集のみで、「長い・あいまい・難しい」の問題点を同時に解決できる。文法によって読みにくい文を判断する手法は、理屈は正しくても実践面で困難さがある。文法の正確な知識が必要になることから、文書作成者自身に負担が掛かり、自分でチェックすることが困難である。百半ルールであれば、文書作成者自身がチェックできる。

また、事物を列記する場合で文が長くなる場合は、必要に応じて、箇条書きを使うこともできる。重要なポイントを文章の中に埋没させないようにするには、箇条書きにするとよい。ただし、文の途中で箇条書きを挟むと、文が分断される。文を完結させてから箇条書きを列挙することが望ましい。

また、箇条書きでは、論理的なつながりを付けずに項目を列挙できてしまう。そのため、十分な説明ができずに、項目の単なる羅列となってしまうことがある。各項目に適宜、必要な説明がされるよう配慮する必要がある。

2 用語集形式 UTX

用語集形式 UTX (Universal Terminology eXchange) は、AAMT (アジア太平洋機械翻訳協会) が策定した、シンプルかつ汎用的で、オープンな用語集形式である (筆者は UTX を策定するチーム リーダーを務める)。UTX は、「仕様」、つまりどのように用語集を作るかという決まりごとである。表 1 に UTX の実例を示す。

表 1 は、AAMT での、「UTX に関する用語」をまとめた用語集から一部を抜き出したものである。左から原語、訳語、品詞、用語ステータスと並んでいる列が UTX の基本的な属性であり、極めてシンプルであることが確認できる。

特許文書では、その分野の専門家なら理解できる用語ならまだしも、文書作成者が、自分にしか理解できない

表 1 UTX 1.11 サンプル (AAMT/UTX 用語集から抜粋)

#UTX 1.11; en-US/ja-JP; 2011-04-19T19:00:00Z+09:00; copyright:AAMT (2011); license:CC-BY 3.0			
#src	tgt	src:pos	term status
Asia-Pacific Association for Machine Translation	アジア太平洋機械翻訳協会	properNoun	approved
dictionary administrator	辞書管理者	noun	approved
Contributor	用語提出者	noun	provisional
Domain	分野	noun	
Glossary	用語集	noun	
Bidirectional	双方向	adjective	approved
Merge	統合する	verb	approved

造語を、きちんと定義することなく、思いつままに無制限に使っていることが多い。文章作成者は、用語が読み手にとって理解しづらいことに気づかない。気づいたとしても、それは読み手の知識や読解力が不足しているからと考え、問題が自分にあるとは自覚しない。ある種の文化として、「理解できないのは読み手が勉強不足である」という書き手優位の考え方が、学校での国語教育から始まって、社会人の作文まで染みついている。

UTX に基づいて用語集を作成し、明確で一貫性のある用語を使えば、「必要な用語は定義して使う」ということがより確実に行える。専門家同士ではある程度の共通理解がある用語でも、文書ごとにニュアンスが異なるので、定義することが望ましい。

造語は、一般的に、無制限に増やすのではなく、必要な場合にのみ作ることが望ましい。造語でないとその概念を表現できないか、慎重に考える必要がある。文書内でしか使われてない新しい概念を表す造語では、定義が

必須となる。このとき、ただ用語を羅列するのではなく、UTX のようなシンプルでありながらも、定まった形式で書くと、検索や翻訳などの各種のツールで用語をデータとして活用できるようになる。UTX では、ツールが使う用語データとしては、定義は必須ではないが、表 2 のように定義を明記すれば、人間の読み手にとっては大いに役立つ。また、対訳を英語で指定することで、より日本語での用語の意味が明確にもなる。用語集という形にまとめれば、複数の用語の相互の関係もより明確にできる。

UTX 用語集は、特許分野でさまざまな活用ができる。特定分野の翻訳では、UTX 形式で、適切な訳語を共有・再利用できる。また、「望ましくない用語」、理解の妨げになる語についての情報も共有できる。

表 2 は UTX 用語集での正規化の例である。

表 2 UTX 用語集での正規化の例

#UTX 1.20; en-US/ja-JP; 2012-04-19T19:00:00Z+09:00; copyright:AAMT (2012); license:CC-BY 3.0				
#src	tgt	src:pos	term status	definition
source language	起点言語	noun	approved	翻訳の原文で使われる言語
source language	ソース言語	noun	non-standard	
source language	原文言語	noun	non-standard	
entry	項目	noun	approved	辞書本文の構成単位
entry	エントリー	noun	forbidden	



UTX の仕様、特に以下に触れる用語ステータス (term status) の詳細については UTX 仕様書を参照されたい。

<<http://www.aamt.info/japanese/utx/>>

ここでは概要を簡単に説明する。

UTX では、辞書管理者と用語提出者という概念がある。「特定の用語集で、ある用語が適切か」ということを、専門知識と権限のあるだれかが責任を持って判断する必要がある。その責任者が、辞書管理者である。辞書管理者が明確であれば、複数の用語提出者からの用語の知識を収集しても、どれが確実に使ってよい用語かを明確にできる。

表 2 では、source language には、「起点言語」と「ソース言語」の 2 つの項目がある。「起点言語」の用語ステータス (表の灰色部分) は approved (承認) であり、これは辞書管理者が承認済みの、正規の訳語であることを示す。「ソース言語」の用語ステータスは non-standard であり、正規の訳語ではないことを示す。

前述の統計的機械翻訳装置の例では、「ソース文」という語が出てくる。これは機械翻訳を専門としておらず、断片的な文章しかない場合では、なにを指しているか読み手に明解に伝わらない。「原文」という用語のほうが明らかに分かりやすい。そして、もし「原文」という語ではニュアンスが不足しているなら、「ソース文」という用語を明確に定義してから使えばよい。

表 2 では、entry という原語については、「項目」という訳語が approved (承認) されているが、「エントリー」という訳語は forbidden (禁止) となっている。これは、この訳語は使ってはならない用語であり、用語チェック ツールなどでのチェック対象となることを示す。

特許明細書に UTX 形式で用語集を添付することが義務付けられれば、誤訳を減らし、より確実に翻訳ができる。

用語集は、文章作成や翻訳で使われているが、さまざまな問題を抱えており、そもそも面倒だからと作られていないことも多い。用語集が存在する場合でも、特定の形式に基づいているのではなく、データとしての形式が一つ一つ異なり、互換性がない。これは複数の用語集を

統合して使いたい場合に大きな障害となる。

必要最低限でありながら実用性の高いシンプルな用語集を作れば、その用語が正しく文書で使われているか、文章作成や翻訳で、ツールを使用してチェックできる。

2012 年 8 月に、UTX について「よくある質問と回答」が公開された。

<<http://www.aamt.info/japanese/utx/faq.htm>>

UTX がどのような用途に適切かなど、具体的な説明がされているので、仕様書とあわせてご覧いただきたい。

言語処理を研究・開発側から見たとき、用語集は「一度作ったらそれっきりの言語データ」と見なされることがある。つまり自分たちで管理して使っていく、という意識がなく、管理するための仕組みがまったく考慮されていない。これは、用語集の作成者と使用者が別の組織 (グループ) であることが一因である。開発側と実際の使用者の間に深い溝があり、十数年以上も埋まる動きがない。UTX では、このような問題に対処するため、用語集を継続的に管理するという観点から、辞書管理者と用語提出者の概念に基づいて、承認・非標準など、実用的な属性に絞って策定されている。

「UTX は単純すぎるのでは」という疑問もよく聞かれる。だが、用語集が必要な企業や組織であるにもかかわらず、長年用語集が作成されずに放置されているのは、用語集の作成や管理が想像以上に複雑だからである。必要最低限のシンプルな情報に限定して、管理の労力を徹底的に減らして初めて、用語集を作成し活用できる。

3 JTF スタイル ガイド

JTF スタイル ガイドは、JTF（日本翻訳連盟）が策定した、翻訳での日本語訳文向けの表記規則である。

英語では、シカゴ マニュアル、MLA、PSA など、学会や企業で、さまざまな場合の細かい表記規則が1000 ページ以上（シカゴ マニュアルの場合）にもわたって規定されている。このような表記規則は、英作文をする人にとっては常識であり、この常識を無視した書き方をすれば、教養のなさを露呈することになる。

一方、日本語では、正書法と呼べるものが存在しない。またコピー エディットの概念も普及していない。つまり、さまざまなカッコや記号を、明確に定義することもなく、乱雑に使っていても苦情はでない。そもそも、このような記号が、明確に論理的な、個別の機能を持つとは認識されてない。いくつか比較的定着している表記規則はあるものの、各業界、各企業独自の表記が雑然として存在している。さらに、企業のエゴにより、企業独自の表記に執拗にこだわる傾向もある。公用文という規則もあるが、現実には厳密に適用されておらず、21 世紀の電子文書を巡る環境の中では複数の問題があり、時代遅れである。

このような状態が問題であるとは、これまで強く認識されてこなかった。そのため、用語集の整備の遅れと同様に、表記の整理の遅れも、企業では目に付く。IT 業界では、厳密で論理的な書き方が求められるプログラミングが関わるためか、比較的、他業界よりも表記や用語への取り組みは行われているが、十分とはいえない。その他の業界では、より状況は悪く、企業ごと、業界ごとにまったく異なる表記が存在する。

表記の不統一は、文章作成、活用、検索、再利用、共有において、費用と労力を増加させるさまざまな問題の原因となる。まず、文書の編集や翻訳の過程で、表記統一に大きな負担が掛かる。また、ウェブ上の文書や組織内のデータベースでも適切な検索ができず、目的の情報を探せない。表記規則がなく、表記が混乱していると、用語集もまたばらばらな表記が使われることになり、表

記の混乱が悪循環に陥る。さらに、表記が統一されていないと、自動翻訳の精度も著しく低下する。ある用語が同じ概念を示しているのか、異なる概念を示しているかが不明確になるからである。

日本語の表記を整理することは、電子文書を効果的に活用するうえで非常に重要である。DTP では、フォントの種類や字詰めには多大な労力が払われる一方、文章の表記（さらには表記によって改善できる読みやすさ）には注意が払われないことも多い。翻訳での表記を整理するため、JTF では、2011 年に翻訳用のスタイルガイドを作成し、公開した。詳細は < http://www.jtf.jp/jp/style_guide/styleguide_top.html > で紹介されており、スタイル ガイドもここから PDF 形式でダウンロードできる。なお、本記事の著者は、JTF スタイル ガイド委員会の委員である。

以下に、JTF スタイル ガイドの 12 のルールを示す。表記の具体例を含む PDF は < http://www.jtf.jp/jp/style_guide/pdf/jtf_style_guide_rule12.pdf > から入手できる。「12 のルール」は、JTF スタイル ガイドでの重要なポイントを簡潔にまとめたものである。

特許分野では、工業分野での表記習慣が反映されていることも多い。工業分野の表記習慣は、必ずしも一貫して明確な合理的理由に基づいておらず、慣習に基づいていることもある。工業分野の表記習慣は、いくつかの点で、JTF スタイル ガイドの 12 のルールと衝突する点がある。たとえば、JTF スタイル ガイドのルール5「長音は省略しない」と、工業分野でよく行われる「長音の省略」が矛盾する。また、ルール6「複合語は中黒または半角スペースで区切る」と、工業分野でよく行われる「区切りなし」が矛盾する。

長音の省略については、JTF スタイル ガイドでは理由は特に挙げていない。ただ、「一般的な文書」では、長音を省略することはないし、長音をそもそも省略することによる利点は特にない。もし「語句や文がより短く書けるから」という理由であれば、英数字が全角で表記されることにより語句や文が長くなる弊害のほうが大きい。

JTF スタイル ガイド 12 のルール

1. 本文を、敬体（ですます調）あるいは常体（である調）のどちらかに統一する。
2. 句読点は「、」と「。」を使う。
3. 常用漢字表にある漢字を主に使用する。
4. 動詞の送りがなは本則に従う。
5. カタカナ語の語尾の長音は省略しない。
6. 長いカタカナ複合語は中黒または半角スペースで区切る。
7. 漢字、ひらがな、カタカナは全角で表記する。
8. 数字とアルファベットは半角で表記する。
9. 原則として記号類は全角で表記する。
10. 半角文字と全角文字の間に半角スペースを入れない。
11. ピリオド（.）、カンマ（,）、スペースは半角で表記する。
12. 単位の表記を統一する。

ルール6の複合語の区切りについては、JTF スタイル ガイド委員会では詳細な検討が行われた。複合語を区切れば、複合語内の構成要素の語が明解になるほか、自動翻訳や音声認識などの自然言語処理に使われる形態素解析でも誤りを防ぐことができるからである。また、語句の検索精度も向上できる。その理由はスタイル ガイド PDF で詳しく説明されている。前述のリンクからスタイル ガイド PDF を入手してご覧いただきたい。

「JTF スタイル ガイドがこのような規則にしたのは、マイクロソフトでの表記変更によるもの」という誤解がたまにあるが、マイクロソフトはまったく関係がない。JTF スタイル ガイドは、合理的な理由と、さまざまな企業での事例や実情を踏まえたうえで策定されているものである。

JTF スタイル ガイドはあくまでも英語など他言語から日本語への翻訳用であることに注意されたい。表記規則としては、文書の執筆用に流用できる点も多いが、日本語での執筆は JTF スタイル ガイドの当初の目的ではない。また文レベルや文章レベルでの「分かりやすさ」を向上することも目的ではない。そのような目的では、当初から日本語作文での「分かりやすさ」の改善を目的とする実務日本語のほうが適しているといえる。

JTF スタイル ガイドは、産業日本語の表記の参考として活用できる。12 のルールに限定しても、そのすべ

てを特許明細書で採用するのは困難かもしれない。ただ、なにも表記基準に基づかない場合や、あいまいで合理的根拠のない「前例」に従って表記するよりは、部分的でも12のルール、あるいは類似のルールに沿うことで、表記を統一できる。1 つでも多くのルールに沿って文書を作成すれば、それだけ手間を減らすことができる。

また、JTF スタイル ガイドについては、現在、3 つのスタイル チェック ツールが以下で公開されている。

<http://www.jtf.jp/jp/style_guide/stylechecktool.html>

作文の規則については、人手によるチェックでは限界があり、必ずチェックのれが発生する。また有償の環境のみでしかチェックできないのも不便である。このようなチェック ツールを公開することは、現実に作文規則を活用してもらううえで欠かせない。チェック ツールの環境がなければ、どんな規則も実際に守られているか、現実的な確認手段がない。

もちろんすべての規則をツールでチェックできるわけではない。作文規則そのものを、チェック ツールでチェックできる規則と、そうでない規則を明確に区別して決定する必要がある。前述の百半ルールは、ツールによるチェックが可能な規則の例である。

特許文書の文体が基本的に常体（だ・である）であることも、作文時に文章を必要以上に難解にする一因と思

われる。常体では、必要もないのに、もったいぶった四字熟語や難しい漢語やカタカナ語をつい使いたくなる。文体を敬体（です・ます）にすれば、読み手を尊重し、より分かりやすく書こうとする意識が高まる可能性もある。

今後、クラウド上での、文書のオンライン共同作成環境が整備されていく。複数の書き手が1つの文書をリアルタイムで作成することがより一般的になると思われる。このような場合でも、複数の書き手がそれぞれ別の表記や用語を使ってしまうと、後で編集や再利用するときに修正の手間が増える。未統一の表記を放置すると、読み手が誤解することもある。最初から共通の表記基準に基づいて文書作成をすれば、最終的には編集での調整をしなくとも、読みやすく、検索・活用・共有・再利用がしやすくなる。

特許分野での表記統一への道のりは長く、険しい。まずは表記についてどのような問題があるかを理解し、表記統一でどのような利点があるかを実感することが手始めとなるであろう、