

明晰特許文書のための 明晰日本語

東京工科大学メディア学部教授

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。

✉ yokoi@media.teu.ac.jp

☎ 0426-37-2435

1 なぜ明晰日本語なのか

日本の産業技術活動は、日本語の上に成り立っている。日本語による情報の作成・蓄積・利用の高度化・効率化は、産業技術活動の高度化・効率化に直接寄与する。そして、分野横断的な情報共有・情報協創を促進させるために、分野ごとに閉じた日本語から他分野へと広く開かれた日本語が求められている。情報のグローバル化やWebグローバル環境を進展させるために、日本に閉じた日本語から世界に開かれた日本語が求められている。コンピュータによる協調支援環境を高度化させるために、人に閉じた日本語からICTへと開かれた日本語が求められている。明晰日本語は、この開かれた日本語である。

特許文書は、産業技術文書の代表例である。特許文書には、日本語を客観的に厳密な情報表現のツールとして使うノウハウが集積されている。また、日々変化する大規模な情報を体系化する情報整理のノウハウが集積されている。しかしながら、特許特有の経緯から特許文書は明晰な文書であるとは言い難い側面を持つ。特許における日本語の用法は、明晰日本語であるとは言い難い側面を持つ。

特許文書の非明晰性は、コンピュータによる文書処理（翻訳、検索など）の精度を下げる大きな要因となっている。これまでは、システム側の工夫による精度向上を

目指してきたが、さらなる精度向上のためには、特許文書の日本語そのものを見直し、明晰な日本語の用法を確立することが重要であることが明らかになってきた。

さらに、特許をめぐるのは、日本語の明晰化を求める大きな動きがある。今日の本格化するプロパテント時代、日本の技術戦略の土台は知財戦略となる時代のもと、技術の研究開発に携わるものは誰もが特許に関わることを求められるようになってきている。誰もが、特許文書を身近にする時代の到来である。さらに、知財ビジネスのグローバル化、特許制度のグローバル化のもと、日本の特許情報に日本語の状態で見ることが世界からアクセスがなされるようになってきている。

2 特許文書に求められる明晰性

特許文書に求められる明晰性には、以下のような側面をあげることができる。これらの側面は、互に依存しあう。なお、コンピュータに対して明晰であるとは、コンピュータが高精度の自動処理や高度な処理支援を実現することができるということである。

① 誰・何に対する明晰性なのか

人に対する

- ・特許の専門家に対する
- ・当該技術分野の通常の専門家（その発明の属する

技術分野における通常の知識を有する者) に対する

コンピュータに対する

- ・ 翻訳処理に対する
- ・ 検索処理に対する
- ・ これからの文書処理に対する

② 文書のどの部分に対する明晰性なのか

出願文書に対する

- ・ [要約] 部分に対する
- ・ [特許請求の範囲] 部分に対する
- ・ [発明の詳細な説明] 部分に対する

審査過程の文書に対する

③ 何時の文書に対する明晰性なのか

既に蓄積されている文書に対する

- ・ 電子出願以前のものを含む
- ・ 電子出願以降のもの

新規に出願される文書に対する

上記の明晰性の各側面に対して、検討すべき明晰性は、以下ようになる。

(1) 人に対する明晰性

人に対しては、共有知識に基づく共通理解を前提とした明晰性を考えればよい。特許の専門家に対しては、構文的な複雑さには慣れていることから、曖昧さのなさや正確な表現という観点から明晰性を検討することになる。一方、当該技術分野の通常の専門家に対しては、一般の技術文書と同等レベルの構文的平易さを実現することを含め明晰性の検討を進めることになる。

(2) コンピュータに対する明晰性

コンピュータの意味理解能力は、ごく一部に留まることから、構文的な制約からの明晰性を検討することになる。しかし、構文構造は、意味構造をガイドする役割を持つことから、構文的な制約に関しても、意味構造の観

点を踏まえて検討を進めることが必要である。なお、簡単な表現構成の繰り返しであれば、人よりコンピュータの方が、より複雑な表現を扱うことができる。

翻訳処理では、文レベルの表現構造は正確に把握できる明晰性が求められる。そして、対象言語の表現特性との親和性という観点からの明晰性が求められる。例えば、英語の文型に対応するように日本語を用いるという観点からの明晰性が求められる。検索処理では、文や文章全体の表現構造を正確に把握するという明晰性は必要ではない。広範な特許文書に対して、簡単な表現要素を網羅的に把握するという明晰性が求められる。簡単な表現要素としては、現状は「語」が中心である。今後は、表現要素を「句」や「文」へと拡大するための明晰性の検討が求められる。また、これからの文書処理をも視野に入れると、意味処理をもう少し陽に加味した明晰性の検討が必要になる。

(3) 特許文書の各部に対する明晰性

現在の特許情報サービスにおいて、網羅的な検索処理・翻訳処理の対象となるのは、[要約]部分である。法的に権利を主張する部分として、最も注意深く著作されるのが[特許請求の範囲]である。そして、この[特許請求の範囲]を具体的に裏付けるのが[発明の詳細な説明]である。

[要約]部分に対しては、通常技術の専門家に対する明晰性と検索処理・翻訳処理に対する明晰性が求められる。[特許請求の範囲]に対しては、権利書としての特許を代表する部分であるから、法的文書としての厳密さと効果的な権利を確保するための意図とを共に反映する明晰性が求められる。それは、特許の専門家に対する明晰性に一致する。[発明の詳細な説明]に対しては、技術文献として特許文書を代表する部分であるから、一般の技術文書と同じ明晰性が求められる。当該技術分野の通常の専門家に対する明晰性である。

審査過程の文書については、日本語の用法がより定型化されていること、著作に関わる専門家が一層限定されること、これらを前提に明晰性の検討を進める必要がある。

(4)各時期の特許文書に対する明晰性

既に蓄積されている特許文書に対しては、検索処理・翻訳処理の観点から[要約]部分、ないし、それに該当する部分を対象に明晰性の検討が行われる。現状の検索処理技術や翻訳処理技術にとって、既存の[要約]部分は十分に明晰であるとは言い難い。そこで、明晰さを補うためにアノテーション付けや書き換えの仕組みを設けることになる。アノテーションに関しては、人にとって十分に軽便・簡明であること、コンピュータにとっては完全な自動処理が可能であることという明晰さが必要である。新たに出願される特許文書に関しては、人にとってもコンピュータにとっても十分な明晰性を確保することになる。ただし、新しい明晰性の導入は、培われてきた特許文化と従事する方々のノウハウが持つ大きな慣性に添う形で進めることが必須要件となる。

3

目標とすべき特許版明晰日本語

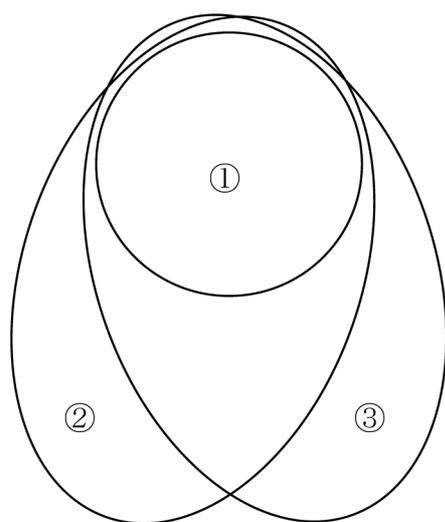
明晰日本語は、一つの日本語仕様を定めようというものではない。それぞれの目的にそれぞれの明晰日本語が必要である。特許版明晰日本語に関しても同様である。共通に満たすべき要件という形で共通の枠組み仕様を策定し、この枠組み仕様を次の3つの具体的な特許版明晰日本語仕様へと具体化する。

① 特許ライティング日本語

明晰な特許文書を作成するための特許用日本語テクニカルライティング仕様である。当該技術分野の通常の専門家が特許文書を一般的な技術文書として読みこなせるようにする明晰性を実現する。明晰性に関して、現在の特許文書は、一般の技術文書が見習うべき点多々ある。これらの見習うべき点は、体系的に整理し仕様に反映する。明晰化の対象となるのは、特許明細書等全体である。

② 日英機械翻訳用日本語

現在特許文書の翻訳に用いられている日英機械翻訳システムの入力テキストを記述するための仕様である。こ



- ①特許ライティング日本語
- ②日英機械翻訳用日本語
- ③次世代文書処理用日本語

図1 特許版明晰日本語（第1次）

の日本語仕様で記述された和文テキストをほぼ意味的に等価な英文テキストに、現状の日英機械翻訳システムがほぼ完全自動で翻訳することができる。明晰化の対象となるのは、[要約]部分を中心に明細書等全体である。

③ 次世代文書処理用日本語

文書処理から知識処理へとグレードアップするために、次世代の特許文書処理を志向するための日本語仕様である。ISeCが開発を進めている概念記述言語(CDL.jpj: Concept Description Language for Japanese)との間で相互変換が可能となる仕様である。明晰化の対象は、特許明細書等全体である。

以上の3つの特許版明晰日本語仕様の相互の関係を図1に示す。①は人に対する明晰性であり、②と③はコンピュータに対する明晰性である。人に対する明晰性は、共通となる意味理解を前提にすることができる。しかし、コンピュータに対する明晰性においては、人であれば意味理解に委ねることができるものも表層表現として明示化することが求められる。②と③は、それぞれ互いに重なり合う部分と重ならない部分とからなる。②の重ならない部分は、英語の表現特性に対応させるために必要となる表層表現の明示化部分である。③の重ならない部分は、概念記述言語(CDL.jpj)の概念記述モデルの特性に対応させるために必要となる表層表現の明示化部分である。

4

仕様の策定作業

Japioの特許情報研究所に設けられた特許版・明晰日本語策定委員会において、枠組み仕様となる基本仕様の策定作業が進められている。作業指針として、現状の特許文書が持つ以下のような非明晰な側面を解消すること

を当面の課題として進めている。すなわち、

- ・一文表記などの、長文による表記
- ・複雑な複合語や複合句
- ・明示化方法の不統一
- ・過剰な明示化による煩雑さ
- ・自然言語（日本語）だけでは記述し難い複雑な構造物や複雑な工程等の表現

また、仕様設計のために以下のような素材を活用する。

- ・特許文書のための日本語ライティング
- ・日本語テクニカルライティング
- ・制限日本語関連
- ・制限英語関連
- ・ワープロソフトの校正・推敲支援機能
- ・文書処理技術における蓄積
- ・言語学の知見

仕様をまとめあげる形式には、④明晰文書を作成するために守るべき書き方ルール、⑤非明晰文書を明晰文書へと書換えていく言い換えルール、この2つが考えられる。従来、この種の仕様は、④の形式にまとめられてきたが、特許版明晰日本語については、⑤の形式、すなわち、言い換えルールの形式にまとめる。その理由は、以下である。

① 自然な作成過程：

特許文書の作成者は、日本語という枠組みや権利記述のための日本語用法の枠組みに従って思考し文書を作成する。これらの枠組みに馴染まない書き方をルール化しても、実効的な作成作業に結びつかない。まずは、これらの枠組みに沿ってすなおに文書を作成し、その文書の部分々々を書き換え、求められる明晰性を達成していくという手順の方が、本来の自然な作成過程である。

② 判断できるルール：

遵守規則としてまとめられた書き方ルールは、実際

の適用に幅がありすぎ、判断が大きく振れることになる。一方、言換えルールにおいては、日本語による言い換えであるから、適用結果が適切であるかどうかを誰もが一般的な基準で判断することが出来る。

③ 支援システムの整備：

いずれ文書作成支援システムが用意される。言い換えは、確度の順に従ってシステムが行ってくれる、あるいは、システムが提示した言い換えの候補から選択することになる。言い換えは、ユーザーとシステムがインタラクションする最適のコミュニケーション手段である。

④ 既存文書への対応：

従来の方式に従って文書を作成し、それを書き換えて明晰性を達成するという手順である。蓄積されている既存の特許文書に適用することができる。[要約]や[特許請求の範囲]を対象に明晰化する作業に適用することができる。

⑤ 書き方ルールとしての利用：

言換えルールは、書き方ルールとして用いることもできる。最初から、この書き方ルールに則って文書作成を行えば、明晰化達成への近道となる。いずれ、明晰ワープロを使って、直接明晰日本語で書くようになる。

5

明晰日本語システム

明晰特許文書の作成に明晰日本語がどのように使われるのか、その概略を説明する。明晰ワープロを中心に置いた明晰日本語システムを図2.に示す。

明晰ワープロは、言い換えルールを適用し、非明晰特許文書を明晰特許文書に変換する作業を支援する機能を提供する。日本語から日本語への翻訳である。通常のワープロと同様にインタラクティブに利用され、言い換の妥当性や他の言い換え候補の選択などの判断はユーザ

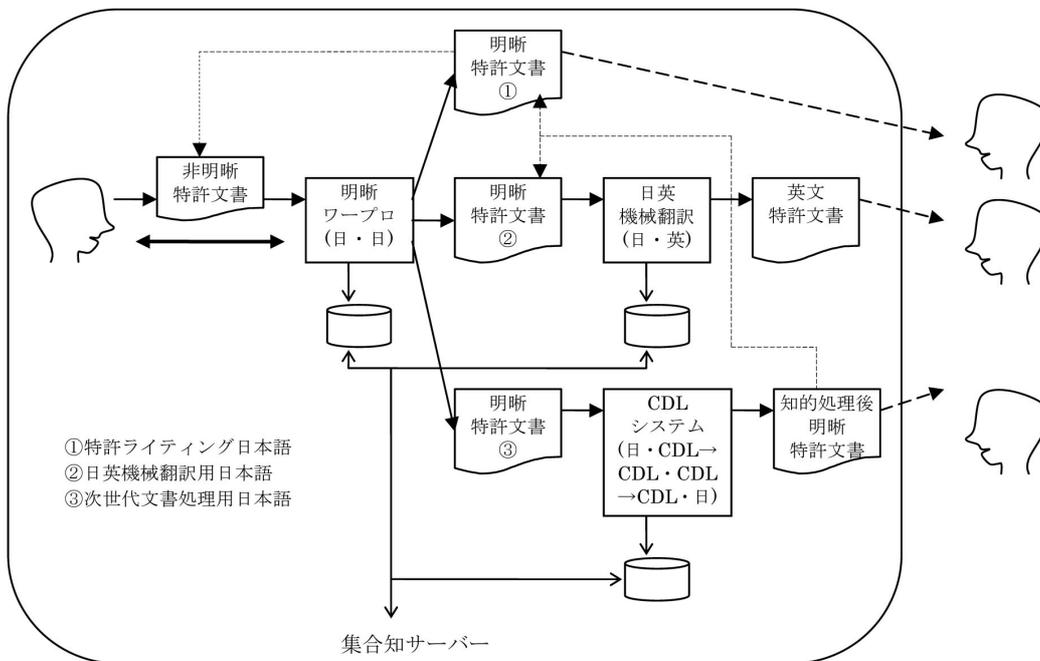


図2 明晰日本語システム

ーが行う。3つのモードを持ち、①特許ライティング日本語、②日英機械翻訳用日本語、③次世代文書処理用日本語、それぞれに対応する明晰特許文書の作成を支援してくれる。

日英機械翻訳は、現在、特許文書翻訳に使われている機械翻訳システムである。日本語から英語への翻訳である。明晰特許文書②の内容を正確に伝えることの出来る英文特許文書が生成される。

CDLシステムは、日本語文書をCDL (CDL.jp) による概念記述に変換し、この概念記述を推論処理し、処理された概念記述から日本語文書を生成するシステムである。日本語からCDLへ、CDLからCDLへ、そして、CDLから日本語への翻訳である。求められる知的処理に従って、さまざまな日本語文書が生成される。明晰特許文書①や②に対応する日本語文書を生成する知的処理も可能である。

明晰特許文書①を経由することによって明晰特許文書②や③を得るという手順を選ぶことも出来る。このような工程を経た方がかえって効率的な場合もあると思われる。

明晰ワープロは、日本語ワープロ技術と日英・英日機械翻訳技術を組み合わせることによって開発することが出来る。開発されるまでは、明晰ワープロの機能をユーザー自身が代行する。そのためのライティングマニュアルが用意されることになる。

範例となる明晰特許文書の蓄積、言い換えルールの改良や追加、新しい専門用語や複合語の辞書登録、これは利用者達による集合知として収集・利用がなされる。図3の明晰日本語集合知サーバーである。サーバーは、全体として共通なものとコミュニティごとに設けられるものが階層的に構成されることになる。

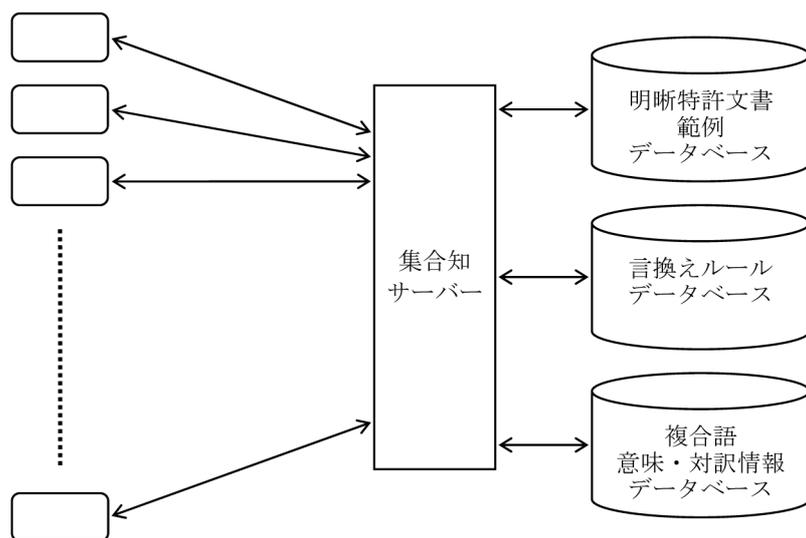


図3 明晰日本語集合知サーバー