

特許文における 複合語の扱いについて

国立情報学研究所
学術コンテンツサービス研究開発センター特任研究員

内山 清子

PROFILE

1997年慶應義塾大学大学院政策・メディア研究科修士課程終了、1998年学術情報センター COE 研究員、2003年スタンフォード大学 CSLI 訪問研究員、2005年学術博士（慶應義塾大学）、2005年～2008年慶應義塾大学特別研究教員准教授、2009年6月より現職、2009年度特許産業日本語委員会委員。



1 はじめに

知的財産戦略の推進に伴い、特許の出願や検索が身近なものになってきている。これから発明する技術については、既存の類似技術の詳細を調べるために特許検索を行い、発明する技術を保護するために出願を行うなど、企業の技術者だけでなく大学の研究者にも「特許」という言葉が浸透してきている。これにより、今までは特定の人にしか読まれなかった特許文書が、より多くの人が読むものとなってきている。また、海外との技術競争に対抗するためには、海外への特許申請も重要となり、他の言語（特に英語）へ翻訳する需要も増えてきている。

従来の特許文書は、知的財産に携わる人にしか理解できないような言い回しや長文により意味を理解することが困難な文が多く含まれていた。しかし、上記のような背景の中で、一般の人にもわかりやすく、文が簡潔で英語にも翻訳しやすいスタイルで書くことが求められている。特に、特許文書には専門用語を代表とする複合語が多く含まれており、複合語の扱いが重要になっている。ここでは、特許文における複合語の扱いについて考慮すべき点や問題点などについて紹介していく。

2 複合語の定義と種類

特許文に使われる複合語には、専門用語と臨時一語が含まれる。まず、複合語とは、自立的な形態素（単語）が二つあるいはそれ以上結合して一つの語を形成するこ

とである。専門用語は、誤りや曖昧さを避けるために、分野の目的に応じて語の意味を限定した用語と定義される。専門用語は分野共通で用いられる用語であり、技術文書、論文、特許文に多く含まれているのが特徴である。さらに専門用語には語彙化したものとそうでないものがある。たとえば、「情報検索」は「情報を検索する」以外の意味は特に含まれていない。一方で「灰上げ機」は「灰を上甲板まで持ち上げ、灰捨て管に捨てる捲上げ機械」を指すため、単に灰をあげるためだけの機械ではない。こうした、専門用語を構成する個々の単語が持つ意味以外の情報を含むことを、語彙化と定義している。

臨時一語は文を凝縮して表現するために、二つ以上の単語を結合して一つの文成分を生成するものであり、即席合成語とも呼ばれる。例として「天然ガス減圧時」「温度低下防止」などが上げられる。臨時一語は専門用語のように固定的に用いられず、説明のためだけに結合して生成されるため、専門用語として広く認識されるものではない。また、構成している単語列から「天然ガスを減圧する時」「温度が低下するのを防止する」のように、元の文に復元可能である。

3 複合語の構成要素

特許文に含まれる複合語は主に複合名詞である。複合名詞を構成する語の種類とその組み合わせを以下に示す。

漢語：名詞、サ変動詞型名詞、形容動詞型名詞

和語：名詞、連用形名詞、形容動詞語幹

・名詞＋サ変動詞型名詞：特許検索

- ・名詞＋名詞：特許文書
- ・形容動詞語幹＋名詞：特殊文字
- ・連用形名詞＋名詞：読み取り装置
- ・形容動詞語幹＋連用形名詞：斜め読み

このように、複合語は複数の語種と品詞の違いにより多様な組み合わせが存在するため、登録されていない複合語の形態素解析および自動抽出を失敗することがあり、扱いが難しいと考えられる。また、上記には含まれない接辞（的、型、用）などが含まれた場合、どこまでを専門用語として抽出するのかの基準設定が困難である。

4 特許文における複合語の問題点

特許文に含まれる複合語は、数が多いことから記述や定義を明確にする必要がある。現時点での問題点として、「専門用語・臨時一語の扱い」「表記の揺れ・不統一」が考えられる。まず、「専門用語・臨時一語」については、従来専門用語は辞書に一括登録することで検索に対応できた。しかし、日々新しい用語が創り出されるため、辞書の更新が間に合わないのが現状である。

次に「表記の揺れ・不統一」については、同分野にも関わらず、他の特許との差異を出すために、新しい複合語を勝手に作る傾向がある。つまり、技術分野に浸透している専門用語とは異なる表記で記述され、用語が統一されていないという現状がある。たとえば「機械翻訳装置」「自動翻訳システム」のように、同じ内容であるが、表記が異なるため検索時どちらかの用語が落ちてしまうことがある。また、海外からの出願においては、単にカタカナに訳した語（例：「原言語」を「ソース言語」など）を新しい技術であるかのように記述してしまうこともある。さらに中黒「構文意味解析」「構文・意味解析」、省略表現「音声認識装置」「音声認識処理装置」などの表記の異なりも検索漏れの原因となるため、早急に対処しなければならない問題である。

上記の問題に加えて、機械翻訳システムなどで自動的に英語に翻訳することを前提とした日本語の特許文については更に注意が必要となる。特に複合語（専門用語）を含む文の場合、サ変名詞が重複して用いられるこ

とが一般的に多い。たとえば、「格納したテキストの翻訳処理を実行する翻訳部」という文章の場合、「翻訳処理を実行する」の部分で「処理」と「実行」が重複している。「翻訳処理」の英訳が日本語と同じ「名詞＋名詞」の「translation process」であれば良いが、サ変名詞の部分をも動名詞で翻訳する場合は「translation processing」となるため、実行する「executing」と重複し「for executing the translation processing of stored text」と訳す場合がある。また、冗長な表現として「訳語選択後の状態において」では「後」と「状態」が重複し、英訳が難しくなる。こうした問題点を今後解消していく必要がある。

5 今後の課題

これまで指摘した問題点について、自動的に処理する側ではさまざまな技術や手法により解決を試みているが、もう一方でこれから特許文書を書く側においても、問題点を認識し、解決するために努力することが望まれる。具体的には、複合語処理の見地として、長い複合語は使用しない（長くても構成単語が5つ以内）、新しい複合語生成時には構成単語を用いて定義づけをする（例：文字列を生成する文字列生成）、使用複合語の統一（制限用語）を徹底する必要がある。特許文書が技術分野の論文と同様に、専門用語を適切に使い、独自性を明確に表現し、分野内の読者に対してわかりやすく説明する文書として定着するためには、特許文書作成のための日本語（産業日本語など）の枠組み、基準作りが重要であると考えられる。

参考文献

- 「現代日本語の複合語形成論」 石井正彦、ひつじ書房 2007年
- 「形態論と意味」日英語対照による英語学演習シリーズ2、影山太郎、くろしお出版、1994年
- 「語の仕組みと語形成」英語学モノグラフシリーズ16、伊藤たかね、杉岡洋子、2002年