

特許文における長い名詞句表現の解析の問題について

株式会社富士通研究所 メディア処理システム研究所主管研究員 **潮田 明**

PROFILE

1983年株式会社富士通研究所入社。表面磁気光学効果、空間光変調器、統計自然言語処理、機械翻訳等の研究に従事。マサチューセッツ工科大学修士、カーネギー・メロン大学博士。産業日本語プラットフォーム委員会委員。

✉ ushioda@jp.fujitsu.com

1 はじめに

特許文の特徴として1文が長く、複雑な複合名詞や長い名詞句を多く含むことなどが挙げられる。このような傾向が生まれる要因として、厳密な記述を行うために多くの修飾語句を用いて意味を限定する必要性や、多くの概念同志の意味的な関わりをひとまとめにして記述するために1文内に収めようとする傾向などが考えられる。またそれとは別に、特許請求項に関しては古くから1文で記載する慣習があり、そのため日常遭遇することがままならない程の長い文が出現することがしばしばある。

人間が読むことだけを前提にした場合は、特許文がどれだけ長くて複雑な文であっても、専門家が見て曖昧性のない文であれば特に問題になることはない。しかし、情報検索、情報抽出や機械翻訳などの処理対象として見たときは、上記の特徴は厄介な問題の原因となる。複雑な文構造の問題についてはすでに本誌でも取り上げたが[1]、本稿では長い名詞句表現を解析する際の問題点について考える。

2 「AによるBのC」型名詞句表現

長い名詞句の代表例としてよく挙げられる「AのBのC」（A、B、Cはそれぞれ名詞）型の名詞句については古くから研究が行われているが[2]、この型の表現はその曖昧性から技術文書では避けられる傾向にある。その代わりに「AにおけるBのC」、「Aに関するBのC」

や「AによるBのC」のような、「の」型名詞句よりも意味を限定できる複合辞を用いた表現が使われることが多い。当然のことながら、「ニーチェの解説書」に比べれば、「ニーチェに関する解説書」や「ニーチェによる解説書」の方が意味的曖昧性は遥かに少ない。しかし、複合辞を用いた名詞句表現でも、他の名詞句を連体修飾したり、他の名詞句から連体修飾されたりすると、意味的曖昧性のみならず、構文的曖昧性をも生じる場合が多い。ここでは複合辞の例として「による」を取り上げ、「AによるBのC」型名詞句表現の構文のおよび意味的曖昧性の問題について考察する。

たとえば「飲酒運転による死亡事故の増加」という表現の解釈について考えて見る。一見曖昧性など何もないように見えるかも知れないが、係り受け関係を考えた場合2つの解釈が存在する。

- A. 「(飲酒運転による死亡事故)の増加」
- B. 「飲酒運転による(死亡事故の増加)」

Aでは「飲酒運転による」は「死亡事故」にかかっているのに対し、Bでは「飲酒運転による」は「死亡事故の増加」あるいは「増加」にかかっている(図1参照)。

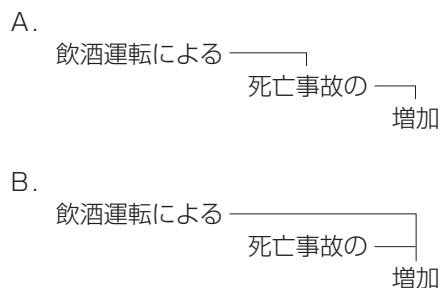


図1 係り受け関係の曖昧性

この構文解釈の違いは意味解釈においても大きな違いを生む。サ変名詞「増加」をサ変動詞で置き換え、サ変名詞「増加」を修飾する連体修飾句を連用修飾句に置き換えることにより、名詞句表現を対応する文に置き換える操作を行うと、A、Bはそれぞれ以下のように書き換えることができる。

A'.「飲酒運転による死亡事故が増加する(している)」

B'.「飲酒運転により死亡事故が増加する(している)」

このような書き換えを行うと、AとBとでは「による」により表されている因果関係に違いがあることがより明確になる。Bは飲酒運転と死亡事故増加の間の因果関係を表している。一方Aは、飲酒運転と死亡事故の因果関係は表しているが、死亡事故(全体)の増加との因果関係は必ずしも表してはいない。たとえば、居眠り運転など他の要因による死亡事故が減少しているために、死亡事故全体の数はむしろ減少しているという可能性も否定はできない。

3 特許文では専門知識なしに解析できない

特許文において因果関係は時に請求範囲の解釈をも左右し兼ねない要素であるだけに、その解釈は重要であるが、情報検索、情報抽出や機械翻訳などの機械処理を行う際にも、名詞句間の係り受け関係の正しい解析は重要である。前節では、係り受け関係の解釈の違いによって、(因果関係という)事実関係に関する解釈の違いが生じる例を見たが、特許文書においてはむしろ事実関係に基づいて係り受け関係を判別しなければならないケースが多い¹⁾。

特許文では、その分野の専門知識や前後の文脈から、係り受け関係の曖昧性は大抵の場合解消可能と考えて良い。逆にそうでない文は特許文として相応しくない文と言える。以下簡単のために、図1 A. のように複合辞「による」が直後の名詞句にかかる係り受けタイプをXタイプ、図1 B. のタイプをYタイプと呼ぶことにする。以

1) もちろん前後の文脈により係り受けの曖昧性が解消されるケースは分野、文種によらず多く見られる。

下の6つの例はX/Yのタイプ判別が比較的容易な例である。

(X) X線束の空気抵抗による減衰の抑制を図り、
(Y) 空燃比の乱れによるエミッションの悪化を防止する

(X) シェルの共振による騒音の低減を図る

(Y) R/Wトランジェントの影響によるスループットの低下を改善する

(X) 圧縮機内の各部において発生する熱による悪影響の抑制を図る

(Y) ポンプの無用な回転によるエネルギーの浪費を防止すると共に、

しかし、特許文においては、以下の例のように専門知識がなければタイプ判別ができないケースも多く存在する。

(X) I/Oバッファ回路の同時スイッチングによる電源ノイズの低減を容易に実施し得る

(Y) 硫黄分の被毒による触媒活性の低下を防止して、

(X) 受光部周辺の光透過層からの光入射によるスミアの防止を図ると共に、

たとえば最後の例では、X解釈では光入射によってスミアが引き起こされるが、Y解釈では逆に光入射によりスミアが防止できることになる。正確な係り受け解析は、用語間の関係や意味内容を用いて検索を行う概念検索においてはすでに重要であるが、自動で技術内容や事実関係を獲得する情報抽出においても必須の条件になってくるであろう。システム開発の立場からは、分野の専門知識を新たに人手で構築することなく、言語処理技術を駆使して今回取り上げた問題等をどこまで解決していけるかが大きな課題となる。

参考資料

[1] 潮田 明: Japio 2009 Year Book (2009) pp.286-289.

[2] 中井慎司、池原悟、白井諭: 「の」型名詞句における品詞情報と意味情報を併用した係り受け規則の自動生成, 情報処理学会研究報告, pp.45-51 (1998.11).