

手がかり句を用いた特許請求項の修辭構造解析

新森昭宏† 奥村学‡ 丸川雄三‡ 岩山真*

shinmori@lr.pi.titech.ac.jp, {oku, maru, iwayama}@pi.titech.ac.jp

†東京工業大学 総合理工学研究科 知能システム科学専攻/インテック・ウェブ・アンド・ゲノム・インフォマティクス

‡東京工業大学 精密工学研究所

*東京工業大学 精密工学研究所/日立製作所

1. はじめに

かつて特許は、機械・製薬・化学など特定分野の企業において、知的財産権担当者や研究者が主に関わるものであった。しかし、ビジネスやサービスの方法を権利の対象とする「ビジネスモデル特許」の出現や、コンピュータプログラムを対象とした「ソフトウェア特許」の認知により、広い範囲の企業関係者が特許に関わらざるを得ない状況が生まれている。大学においても、研究成果や技術移転のコアとしての特許の重要性が認識されるようになった。

特許は一種の法的文書であり、その内容を記述する「特許明細書」は、独特の記述形式を持っている。最も重要な箇所である「特許請求項」(クレーム)は、1文で発明内容を記述するという制約と、独特の記述スタイルにより、専門家以外の人にとっては極めて読みにくいものになっているのが通例である。

特許行政は、いち早く電子化が推進された分野であり、過去に出願された特許明細書の電子データは膨大に蓄積されている。さらに、年間40万件以上と言われる新規出願に伴い、データ量は日々増加している。こうした膨大な特許データを対象とした研究は従来、検索に関するものがほとんどであった。

我々は、自然言語処理の観点から特許データを対象とした研究を進めている。当初、特許請求項の可読性向上のための言い換え[1]という観点から研究を開始したが、現在はNTCIR3ワークショップ(主催:国立情報学研究所)[2]に参加し、特許マップ作成のための基盤技術確立という方向性で研究を行っている。ここで、特許マップとは、関連する複数特許の相互関係分析、特定分野での出願傾向分析などの用途で用いる概念的ツールを意味している。

本稿では、特許マップ作成の基礎技術としての特許請求項の修辭構造解析について、手がかり句を用いた手法を提案し、現時点における実装状況について報告する。

2. 特許請求項の記述特性

2.1 一般的特性

特許請求項は、新聞記事や論文などとは異なり、読みやすさよりも正確性や厳密性が重視される。また、出願人が主張しようとする権利範囲を明確に曖昧性なく記述することになっているため、記述が長くなることが多く、構文も複雑になる。さらに、専門家(弁理士や知的財産権担当者)が作成することが多いため、一種の専門家コミュニティが形成されており、一定の記述スタイルが確立している。

特許出願者の立場から「どのように特許請求項を記述すべきか」を解説した文献[3]では、特許請求項の記述スタイルに関して、表-1に示すように、3つのスタイルへの類型化が行われている。ここで、各類型は排他的なものではない。たとえば、ジェブソン形式の公知部分または新規部分が、書き流し形式または要件列挙形式で記述されることは多い。

表-1 特許請求項の記述スタイルの類型

スタイル	説明
書き流し形式	・「…し、…し、…した、…」のように、処理内容を「書き流す」形で記述する。 ・方法の発明で用いられることが多いが、物の発明において用いられることもある (product-by-process claim)。
要件列挙形式	・「…と、…と、…とからなる、…」のように、構成要素を列挙する形で記述する形式。 ・主に、物の発明で用いられる。
ジェブソン (Jepson) 形式	・「…において、…を特徴とする、…」、「…であって、…を特徴とする、…」のように、最初に、既に知られている内容(公知部分)を述べた上で、この発明の特徴となる部分(新規部分)を記述する形式。

2.2 言語処理の観点から見た特性

(1) 明示的に挿入された改行

長い特許請求項の場合、下例のように、改行が明示的に挿入されていることがある。

<例> (特願平8-182670より引用)

【請求項1】 原稿が載置される原稿台と、改行
この原稿台に対して主走査方向に移動する走査光学
手段と、改行
この走査光学手段上に配置され原稿を副走査方向に
照明する照明手段と、を備えた画像読取装置におい
て、改行
前記照明手段は、前記走査光学手段に対して走査移
動平面に略平行に回動自在に取付けられることを特
徴とする画像読取装置。改行

(注：上記において、改行と明記した箇所に、改行コードが挿入されている。なお、表示レイアウトのためのマークアップ(
)は削除している。)

NTCIR3の「特許データコレクション」(約6万件)を調査した結果、第一請求項の記述中に改行が挿入されているものの割合は、約46%であった。

こうした改行は、特許明細書作成者(弁理士等)が自分で理解しやすいようにするために、記述のまとまりごとに挿入したものと推測される。

言語処理の観点からは、こうした改行の近傍に頻出する表現を、改行が挿入されていない請求項の解析のための手がかり句(cue phrase)として利用し、長い請求項の断片化に利用するということが考えられる。

(2) 記述末尾の「名詞まとめ」直前の修飾表現

特許請求項は通常、名詞句であり、その末尾は「○○の◇◇」の形式をとることが多い。ここで、○○と◇◇は、名詞または複合名詞であるが、特許の場合は、造語や辞書に登録されていない未知語であることが多い。そして、末尾の「○○の◇◇」の直前には、「を特徴とした」のような定型的な修飾表現が付けられていることが多い。こうした表現も手がかり句として利用できる。

(3) 請求項解析のための手がかり句

以上の予備調査結果と、表-1のような類型化を考慮した目視分析等により、表-2に示すような手がかり句を収集した。(なお、表-2では、手がかり句の表記に、Perlの正規表現記法を使用している。)

表-2 特許請求項解析のための手がかり句

出現箇所	手がかり句
ジェブソン形式で、公知部分と新規部分、前提部分と本論部分を区切る箇所	に(お 於)いて、 であって、 にあたり、 に当(た)?り、
特徴を明示する箇所	を特徴と(した する)(、)?
構成を明示する箇所	を(具 備 そな)え(た る)(、)? を具備(した する)(、)? で構成され(た ている)(、)? を有(する した)(、)? を包含(する した)(、)? を(含 む んだ)(、)?
要件列挙形式で、構成要素(名詞句)を並列させる箇所	(名詞)と、
書き流し形式で、処理(述語)を並列させる箇所	(動詞連用形)、

3. 修辞構造理論の特許請求項解析への応用

日本語の長い文の構造解析において、シソーラスとダイナミックプログラミングを用いて並列構造を検出するアルゴリズムと、それに基づく係り受け解析プログラムが開発され、科学文献や新聞記事の解析において一定の精度が達成された[4]。

しかし、特許請求項の記述は複雑であり、かつ、1つの事項を説明した後で、それを用いて別の事項を説明するという、連鎖的な記述が多くみられるため、このアルゴリズムが必ずしもうまく動作しない。たとえば、前述例の「特願平8-182670」の場合、「～原稿台と」・「～走査光学手段と」・「～照明手段と」の3つの部分を並列構造としなければならないが、下線・二重下線で示すような連鎖の存在により、この並列構造がうまく検出されない。

また、特許マップ作成という最終目的から考えると、真に必要なのは係り受け解析結果ではなく、記述されている構成要素とその関係であるとも言える。

そこで我々は、特許請求項の解析のために、修辞構造理論(RST: Rhetorical Structure Theory) [5]を適用することを考えた。修辞構造理論は、自然言語の談話構造を解析するための理論的枠組みとして

表-3 特許請求項用の修辭関係とその検出方法

区分	修辭関係	説明	検出方法
多核	PROCEDURE	書き流し形式の場合 例: [~し、][~し、][~する]XXX 	「書き流し形式で、処理(述語)を並列させる箇所を示す手がかり句」が存在するとき
	COMPONENT	要件列挙形式の場合 例: [~と、][~と、][~と]を備えた XXX 	「要件列挙形式で、構成要素(名詞句)を並列させる手がかり句」と、「構成を明示する手がかり句」が存在するとき
単核	ELABORATION	詳述 例: [XXX を YYY した][ZZZ の AAA] 	「ZZZ の AAA」等の形式をした「名詞まとまり」が後ろの部分に存在し、その直前には他の修辭関係に組み込まれていない部分が存在するとき
	FEATURE	特徴付け 例: [XXX した][ことを特徴とする]YYY 	「特徴を明示する手がかり句」が存在するとき
	PRECONDITION	公知部分、または前提部分 例: [XXX であって、][YYY した]ZZZ 	「ジェブソン形式の公知部分と新規部分、前提部分と本論部分を区切る手がかり句」が存在するとき
	COMPOSE	構成対象 例: [~と、~と、~と][を備えた]XXX 	「構成を明示する手がかり句」が存在するとき
	DEFINITION	用語定義 例: [XXX は~で構成され、][この XXX に基づいて、~] 	提題の「は」が存在するとき

1980 年代に提唱され、その後、自動要約[6]や自動レイアウトなどに応用されている。修辭構造を対話型で編集するためのツールとして、Tcl/Tk 言語による RSTTool [7]も開発されている。

特許請求項の修辭構造を解析するにあたり、我々は表-3 に示すような、修辭関係を定義した。また、同じく表-3 に示すように、それぞれについて、手がかり句を用いた検出方法を考案した。なお、表-3 において、多核(multi-nuclear)とは、関係を構成する要素群が対等である関係を指し、単核とは、重要な要素(核: nucleus)と補足的な要素(衛星: satellite)とから構成される関係を指す。単核関係の説明中の矢印は、衛星から核の方向に引かれている。

修辭構造解析を行うことにより、特許請求項の構成要素を明確化できることが期待される。

4. 手がかり句を用いた修辭構造解析プログラム

特許請求項をトップダウンに解析し、表-3 に示したように、手がかり句を検出しながら断片化し、

各手がかり句に応じて、修辭構造木を組み上げる処理を行うプログラム(rst_claim)を実装している。

特許請求項は通常、最初に記述されるものが最も記述が長く、かつ重要であることが多いため、処理の対象をこの第一請求項に絞ることとした。表示レイアウトのためのマークアップ(
)を削除し、記述のまとまりを区切るためにもともと挿入されている改行は保存するようにして、切り出したテキストを入力とした。出力は、RSTTool[7]の v2.7 におけるファイルフォーマットである「.rs2 ファイル」(XML ライクな形式で、修辭構造をマークアップしたファイル)とした。このことにより、rst_claim の実行結果を視覚的に確認することができる。

図-1 に、修辭構造解析プログラム(rst_claim)の位置付けを示す。



図-1 修辭構造解析プログラム(rst_claim)の位置付け

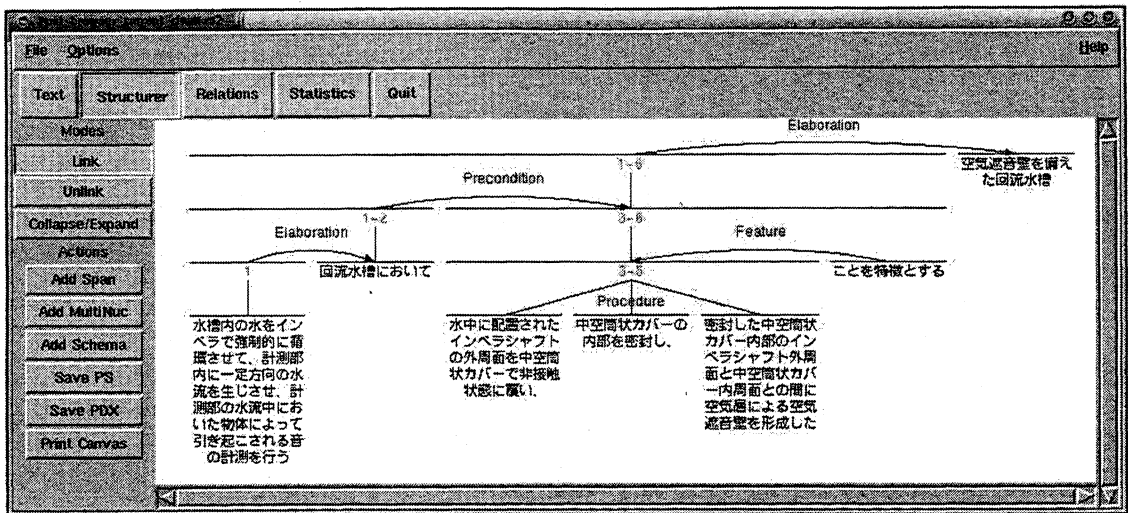


図-2 特許請求項の修辞構造解析結果 (RSTTool v2.7 により表示)

以下のテキストを入力して修辞構造解析を行った結果をRSTToolで表示している様子を図-2に示す。

<入力> (特願平8-181191より引用)

水槽内の水をインペラで強制的に循環させて、計測部内に一定方向の水流を生じさせ、計測部の水流中においた物体によって引き起こされる音の計測を行う回流水槽において、水中に配置されたインペラシャフトの外周面を中空筒状カバーで非接触状態に覆い、中空筒状カバーの内部を密封し、密封した中空筒状カバー内部のインペラシャフト外周面と中空筒状カバー内周面との間に空気層による空気遮音壁を形成したことを特徴とする空気遮音壁を備えた回流水槽。

5. 今後の課題

修辞構造解析プログラムは実装途上であり、現時点ではフォーマルな定量的評価を行う段階にはない。これまでに実施したインフォーマルな実験では、いくつかの良好な結果が得られた。手がかり句だけを利用した表層的な処理アルゴリズムで、ある程度の結果が得られるということは、特許請求項の記述スタイルがそれだけ確立しているということを示していると言える。

現在の実装では、修辞構造解析プログラムは、特許請求項テキストだけを入力しており、「発明の詳細な説明」は参照していない。特許法は、特許請求項の記述内容は必ず「発明の詳細な説明」で説明されなければならない旨の規定をしており、また、「発明の詳細な説明」は、特許請求項記述に比べれば短く、かつ、日常的な文体で記述されることが多い。

このため、「発明の詳細な説明」のテキストまたはその解析データを利用することで、修辞構造解析の精度を向上させることができる可能性がある。

特許請求項の修辞構造を解析することで、特許の構成要素を抽出することができる。これにより、言い換えだけでなく、複数特許間の相互関係分析など特許マップ作成への展望も開けると思われる。

謝辞

本稿及び本研究では、NTCIR3 ワークショップで配布された「特許データコレクション」を使用しています。

参考文献

- [1] 新森昭宏、齋藤豪、奥村学、「特許請求項の可読性向上のための自動言い換えについての考察」、言語処理学会第7回年次大会併設ワークショップ「言い換え/パラフレーズの自動化に向けて」、2001年3月。
- [2] 岩山真、藤井敦、高野明彦、神門典子、「特許コーパスを用いた検索タスクの提案」、情報処理学会情報学基礎研究会 No.063-007, 2001年7月。
- [3] 葛西泰二、「特許明細書のクレーム作成マニュアル」、工業調査会, 1999。
- [4] 黒橋禎夫、「結構やるな、KNP」、情報処理学会誌、Vol.41, No.11, 2000。
- [5] Bill Mann, "An Introduction to Rhetorical Structure Theory (RST)", 1999。
(<http://www.sil.org/~mannb/rst/rintro99.htm>)
- [6] Daniel Marcu, "The Theory and Practice of Discourse Parsing and Summarization", MIT Press, 2000。
- [7] Michael O'Donnell, "RST-Tool: An RST Analysis Tool", Proc. of the 6th European Workshop on Natural Language Generation, 1997. (<http://www.wagsoft.com>)