

特許文請求項からの定型パターンの抽出

Extraction of Typical Patterns from Patent Claims



山形大学名誉教授

横山 晶一

1949年生。1972年東京大学工学部卒。同年電子技術総合研究所入所。1991年同所知能情報部自然言語研究室長。1993年4月より山形大学。2015年4月より同名誉教授。工学博士。アジア太平洋機械翻訳協会（AAMT）/ Japio 特許翻訳研究会委員。

✉ yokoyama@yz.yamagata-u.ac.jp

1 はじめに

特許文の課題、解決手段、請求項の文が、長い修飾語と複雑な文の係り受け構造を持ち、人間にとっても意味が取りにくく、翻訳がしにくいこと、また、機械翻訳でもその品質が今一つであることはよく知られている。

統計的機械翻訳（SMT）においては、請求項の構造を sublanguage ととらえて組み込むことによって、翻訳品質が改善されることが報告されている^[1]が、意味構造を正確にとらえるためには、なお解析が必要であると考えられる。

また、ニューラルネットを用いた機械翻訳（NMT）では、学習に大量の二言語コーパス（bilingual corpus）を必要とし、そこで得られた翻訳結果の意味づけや、訳語の欠如や誤謬、もともと存在しない情報の挿入などを修正するのが困難であるということが指摘されている。

ここ数年来、日本語特許文の請求項 1 に焦点を絞って、並列接続詞、並立助詞、入れ子構造などを調査してきた^[2, 3]。本稿では、その中から、請求項の中の定型文として使用できそうなパターンを抜き出して、その可能性を探る。定型的な文のパターンがある程度分かることによって、そのパターンに近いものをバイアス的に学習すれば、ランダムに大量のデータを学習するのに比べて、機械翻訳である程度改善された結果が出てくる可能性があると考えている。

2 資料・調査方法

これまで調査したことのない特許公報（特開 2003-180626～180738）の中から、請求項 1 の一文が 120 文字を超えるものを 100 選び出した。内訳を表 1 に示す。

表 1 調査特許文請求項 1 の文字数の内訳

文字数	120～ 200	200～ 299	300～ 399	400 以上
特許数	37	34	18	11

以前の調査よりも少し長めのものが多いが、最も短い文は 127 文字、最も長い文は 643 文字、平均は約 262 文字で、大きな傾向としては、従来とそれほど変わってはいない。

調査も、従来の手法を踏襲して、次のような項目について行った。

(a) 機能語「であって、」、「において、」による文の分離

これまでの報告で述べてきたように、特許文の請求項では、「[名詞句 A] [機能語（であって、においてなど）]、…した（する） [名詞句 A]」という形をとるものが非常に多い。しかも、「であって、」や「において、」の部分で、全体を 2 つに区切れる例がほとんどであることが、従来の調査で判明している。これについて、今回も確認した。

(b) 「特徴とする」の調査

上記の名詞句で、最後の部分の直前の「した（する）」が、「特徴とする」、「特徴とした」となっていることが

多いこともすでに報告した。今回もこれに関する調査を行った。

(c) 「と」による並立構造

長い名詞句にかかる修飾句は、名詞を読点で区切って並べる場合も少数見られるが、その多くは、[名詞句]+「と、」という形になる。

(d) 動詞による並列構造

「と、」という助詞ではなく、「～備え、」、「～有する」といった動詞による並列構造も多い。

(e) 改行による係り受けの明確化

主要な修飾句の部分に改行を入れることによって、係り受け関係を明確化した請求項も比較的多く見られる。

(f) 「前記」という記述による前の名詞句の省略

「前記」という名詞は、英語の“the”や“its”などに相当し、前の名詞や名詞句を受ける働きをする。「該」も同じような役割を果たすが、今回は「前記」についてのみ調査を行った。

これらの結果について、次節で述べる。

3 調査結果

(a) 機能語「であって、」、「において、」による文の分離

これまでの報告と同じように、表1に示す100文について、調査した結果を表2に示す。

表2 「であって」と「において」の内訳

	典型	非典型	その他	
			読点あり	読点なし
であって	31	4	1	0
であり	1	0	3	1
において	35	3	2	3

ここで、典型というのは、「[名詞句 A] [機能語(であって、においてなど)]、…した(する) [名詞句 A]」の形になっていて、機能語の部分で句が前後に分割されるものを示し、非典型とは、前後の名詞句が少し異なっているものを示す。その他とは、修飾句のもっと下位にあって、そこでは一つ上の名詞に係る形になっているが、句の大きな分割には寄与しないものを示している。

表2に示すように、「であって」や「において」を含む句の多くは、その部分の前後で句が大きく分割される。100文の約6割が、このような形で句の長さを短くして分析できる。

ただし、句の分割のしかたはさまざまで、必ずしも前後の句が均等に分かれるわけではない。比較的均等に分かれた例を図1に示す。この図では、ほぼ均等に句が分割され、しかもそこに改行も入り、解析しやすくなっている。「特徴とする」という句も語末に入っている(これらについては後述)。ここで示される名詞句も「人体に装着される運動量測定装置」という、非常に長いものとなっている。

生体の体動を検出する加速度センサと、この加速度センサにより得られる信号に基づき生体の運動量を演算する運動量演算手段と、算出された運動量を表示する表示部とを備える人体に装着される運動量測定装置において、
所定期間の推定消費カロリーを用いて残目標カロリーを算出する残目標カロリー算出機能と、生体が所定期間に消費するのに望ましい目標運動量を算出する目標運動量算出機能、又は実測された運動量から生活活動指数を算出する生活活動指数算出機能とを備え、運動量、目標値と実測値との差又は生活活動指数の時間傾向を前記表示部に表示することを特徴とする人体に装着される運動量測定装置。

図1 「において」で分離される典型例(特開2003-180665)

(b) 「特徴とする」の調査

前の報告^[3]にも示したが、「特徴とする」を含む例は非常に多い。今回の調査でも、「特徴とする」が77例、「特徴を有する」が1例、「特長とする」が1例の計79例あり、そのいずれもが、語末の名詞句の直前に置かれたものであった。これと(a)とのかかわりについては、次節で述べるが、請求項に見られる典型的なパターンを形成するものとして重要である。なお、今回は、前回のような「特徴とした」というパターンは見られなかった。図1はその典型的な例となっている。(c)に示す図2もその例である。

(c) 「と」による並立構造

名詞句を列挙して、「～と、…と、」という形で修飾句を並列させる構造も、請求項に特徴的なパターンである。

今回調査した中でも、7割近くが、助詞「と」を繰り返すことによって、修飾句の並列を形成している。図

2は、「において」、「と、」による並立、動詞による並列、「前記」、「特徴とする」を含む典型例であり、改行も含んだ長い句を示している。

浴槽内に設置された電極によって検知された入浴者の心電信号に基づいて、所定の信号処理を行う心電信号処理装置において、
前記電極によって検出された心電信号を増幅する増幅手段と、
前記増幅手段によって増幅された心電信号に基づいて所定の心電情報を作成する心電情報作成機能を有する制御手段と、
前記浴槽に貯留された湯水の導電率を測定する導電率測定手段とを備え、
前記増幅手段は、前記浴槽に貯留された湯水の導電率に応じて、入力される心電信号の増幅率を変更可能に構成されていることを特徴とする心電信号処理装置。

図2 並立助詞等を含む例（特開 2003-180643）

(d) 動詞による並列構造

動詞は、以前の報告で述べた「備える」や「有する」といった例（活用して、「備え」、「備えた」、「有し」といった形で（読点とともに）用いられる場合も多い）もあるが、サ変動詞の並列など、さまざまなバリエーションがある。今回は、あまり動詞の形を特定せずに、動詞一般として調査を行った。全体の約半数に、この形が見られる。つまり、おおざっぱに言って、助詞、動詞両方の並列的な構造を含むものが2割近くを占めていると言える。図2もその一例である。

(e) 改行による係り受けの明確化

請求項1の長い句の中に改行を入れている例は比較的多く、今回も64例見られた。図1のように、主要な区切りのところに1か所だけ改行を入れているものと、図2のように、修飾句ごとに区切りを入れているものがあるが、全体的には後者の方が多い。数は少ないが、あまり的確でない場所に改行を入れ、かえって解析の妨げになるものも見られる。

(f) 「前記」という記述による前の名詞句の省略

「前記」という語によって、前の句とのつながりが明確になる場合が多い。今回は、約半数の51例でこの記述が見られた。図2は比較的明確な例である。「電極」、「増幅手段」、「浴槽」といった語句が、前とのつながりを明確化する形で記述されている。ただし、修飾句のどのレベルでこれが使われているかは、句内部の意味記述を明らかにしないと解析できない場合も多く、今後の課題である。

(g) その他

「前記」と似ているが、「該」という語を含む例も比較的多く見られた。この語は、「前記」よりも曖昧な使われ方をする場合が多く、今後さらに検討が必要である。

箇条書きも、しばしば見られるが、今回調査した範囲には2、3例しかなく、十分な知見が得られていない。

また、「とともに」（「と共に」と記述される場合もある）などの機能語についても、今回は十分には調査していない。今後これらにも手を広げる予定である。

4 典型的なパターンの抽出と調査

前節に述べたいくつかのパターンを組み合わせると、請求項に特有のパターンが浮かび上がってくる。

表3は、「であって」、「において」と「特徴とする」を組み合わせた時の結果を示している。一番左のカラムは、

(a) [名詞句 A] であって、～を特徴とする [名詞句 A] を示し、左から3番目のカラムは、

(b) [名詞句 A] において、～を特徴とする [名詞句 A] を示している。(a)には、「特徴とする」の代わりに「特徴を有する」という形が1、(b)には「特長とする」の1を加えてある。(a)の例を図3に、(b)の例は、図1、2に示してある。

表3 典型的なパターンのまとめ

であって		において		であり		その他
典	非	典	非	典	非	
20	3	32	2	1	0	21

わずか100の例を調査しただけで、全体像を示すのはやや乱暴であるが、請求項の約半数が、この形のパターンを持っていることが期待できる。



湯槽 (2) と、この湯槽 (2) に貯留した湯 (H) を加熱・保温するためのヒーター (3) と、前記ヒーター (3) の温度調整装置 (4) と、複数の孔 (5b) を貫通形成した足載置板 (5a) を有する足置台 (5) とを備え、前記ヒーター (3) 及び前記足置台 (5) を前記湯槽 (2) 内に貯留した湯 (H) に浸漬させて使用する足温器であって、遠赤外線効果を有するセラミック粒 (11) をさらに備え、このセラミック粒 (11) を前記湯槽 (2) 内に貯留した湯 (H) に浸漬させて使用することを特徴とする足温器。

図3 典型的なパターンの一例 (特開 2003-180724)

このような機能語がない例を図4に示す。「特徴とする」は、ここで調査した限りでは、必ず最後の名詞句の直前に位置しているので、機能語が無い場合でも、大きな構造を捉えるのに役立つ可能性がある。

内視鏡観察時に照明用光源として用いられるランプと、前記ランプに電力を供給するランプ電源部と、前記ランプに印加されるランプ電圧に基づいて前記ランプの残余使用可能期間に対応するランプ余命を算出するランプ余命算出手段とを備えることを特徴とする内視鏡装置のランプ余命算出装置。

図4 機能語を含まない例 (特開 2003-180632)

5 おわりに

NMT 全盛の時代に、文の構造を調査するのは、ある意味無駄な作業に思えるかもしれない。しかしながら、照応などの一文を超える現象などに関しては (最近では簡単な照応現象をニューラルネットで学習する試みも行われているが)、まだ、ある程度の基本調査が必要と考えられる。

理想的には、たとえば、昔話の解析のように、いくつかの定型的なあらすじをさらに一般化するところまで広がって行ければ面白い^[4] が、特許文の請求項に対してこのような試みが成功するかどうかは不明である。

今後は、さらにデータを増加させるとともに、ここでは触れられなかった諸現象にも眼を配り、できる限り談話的な部分を含んだ一般化を進めていきたいと考えている。また、さまざまな現象がどのように組み合わせさせて

いるかについてもさらに研究を進めたい。

また、系統的に調査したわけではないが、アメリカやドイツの特許の請求項 (英: claim、独: Anspruch) もやはり長い名詞句であることが多いようである (たとえば、ドイツ語では、関係副詞 “wobei” (英語 wherein、その場合) などを使って、長々とした名詞句にする例などが見られる)。今後これら外国特許とも比較してみたい。

参考文献

- [1] 富士秀、藤田篤、内山将夫、隅田英一郎、松本裕治：特許請求項に特有の文構造に基づく英中日特許請求項翻訳、自然言語処理 Vol.23, No.5 (2016) pp.407-435
- [2] 横山晶一：パターンによる特許文請求項の構造解析、Japio YEAR BOOK 2017 pp.298-301
http://www.japio.or.jp/00yearbook/files/2017book/17_4_06.pdf
- [3] 横山晶一：パターンを用いた特許文請求項の構造解析、平成 29 年度 AAMT/Japio 特許翻訳研究会報告書 (2018) pp.32-37
http://aamtjapio.com/kenkyu/files/kenkyu05/AAMT_Japio_20180423.pdf
- [4] マックス・リュティ、小澤俊夫訳：ヨーロッパの昔話、岩波文庫 (2017)