

# 特許請求項定型文の分類と解析

Classification and Analysis of Typical Patterns in 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

特許文の請求項が長文で複雑な構造を持ち、その正しい意味を理解するのが難しいということは、すでに言及してきた<sup>[1]</sup>。また、その中に決まったパターンがあり、米国やヨーロッパの特許においても、同様のことが言えることも指摘してきた。

図1 (a), (b) は、パテントファミリーをなす特許の英語とドイツ語の請求項1の例を示したものである。

A semiconductor component **comprising** at least one field effect transistor, said transistor **comprising** at least a back barrier layer, a buried layer arranged on the back barrier layer, a channel layer arranged on the buried layer, a barrier layer arranged on the channel layer, and a gate layer arranged on the barrier layer, **wherein** the barrier layer comprises Al<sub>2</sub>Ga<sub>1-z</sub>N, **wherein** the buried layer comprises Al<sub>x</sub>Ga<sub>1-x</sub>N and at least one dopant causing a p-type conductivity, and the gate layer comprises any of GaN and/or Al<sub>u</sub>ln<sub>v</sub>Ga<sub>1-v</sub>uN.

図1 (a) 米国特許の請求項の例 (US20170373177A1)

Halbleiterbauelement, **welches** zumindest einen Feldeffekttransistor (1) **enthält**, **welcher** eine Rückseitenbarrierschicht (2), eine darauf angeordneten Einbettungsschicht (3), eine darauf angeordnete Kanalschicht (4), eine darauf angeordnete Barrierschicht (5) und eine darauf angeordnete Deckschicht (6) **enthält**, **wobei** die Barrierschicht (5) Al<sub>2</sub>Ga<sub>1-z</sub>N enthält oder daraus besteht, **dadurch gekennzeichnet, dass** die Einbettungsschicht (3) Al<sub>x</sub>Ga<sub>1-x</sub>N und zumindest einen Dotierstoff enthält oder daraus besteht, **wobei** der Dotierstoff eine p-Leitfähigkeit bewirkt und die Deckschicht (6) GaN und/oder Al<sub>u</sub>ln<sub>v</sub>Ga<sub>1-v</sub>uN enthält oder daraus besteht.

図1 (b) ドイツ特許の請求項の例 (DE102017210711A1)

(a), (b) はほぼ同じ内容（半導体部品）を示しており、構文的にも似通っている。(a) で、“comprising”（含む）という分詞形でつないでいる部分は、ドイツ語では関係代名詞“welches”（中性名詞主格を示す）と、動詞“enthält”（“enthalten”（含む）の3人称単数形）で記述されているし、後半の英語の関係副詞“wherein”は、ドイツ語では、“wobei”が対応している。英語とは必ずしも対応していないが、ドイツ語の“dadurch gekennzeichnet, dass”（～によって特徴づけられる）という表現も、ドイツ語特許の請求項にはよく見られるパターンである。日本語特許にも、類似のパターンが見られることは、すでに述べた。本稿では、これまでと同様に、これら請求項に見られるパターンを引き続き調査するとともに、これまで行ってきた調査結果と合わせて、全体像を明らかにする手掛かりとしたい。

自然言語処理や機械翻訳の分野では、ニューラルネットを用いた手法が主流であり、特に最近のTransformer や、それを活用したBERTなどは、すでにいろいろ活用され、今後の応用も期待されている（このあたりの動向については、[2] が詳しい）。

人手で調査した少数の結果程度では、大量の学習データを要するニューラルネット手法にはなじまないという意見もあるが、学習済みのネットを別の少数のラベル付きデータに適用して精度を高めようというファインチューニングの技法なども提案されており<sup>[3]</sup>、この調査データに基づいてラベル付けしたデータを、これらの応用に用いることも考えられる。

## 2 資料・調査方法

例年と同様に、2003年の特許公開情報の中（特開2003-180881～180994）から、請求項1が120文字以上のものを100文選び出した。表1にその内訳を示す。表1には、以前に行った450文（特開2003-180203～180880）の内訳も併せて示す。

今回調査した100文は、最も短いものが123文字、最も長いものが975文字（これは異例に長いといえる）、平均で約273文字であった。今回調査した部分は、遊技機に関する特許が比較的多く、文もやや長い傾向が見られたが、おおむね従来行った450文に関するデータの内訳とそれほど変わってはいない。

表1 特許文請求項1の文字数の内訳

文字数		120～199	200～299	300～399	400以上
特許数	前回まで	200	151	62	37
	今回	32	34	21	13
	合計	232	185	83	50

調査項目は従来のものと同じである。次節の調査結果において、項目内容と一緒に示す。

## 3 調査結果

### (a) 機能語「であって、」、「において、」による文の分離

請求項に含まれる文での典型とは、

・[名詞句A][機能語]…する(した)[名詞句A]  
という形のものである。[機能語]の部分には、いろいろな句が入りうるが、今回は、「であって、」と「において、」に限定した。前回の報告<sup>[1]</sup>では、「であり、」も含まれていたが、数が少なく（今回対象の100文についても5文あり、そのうちの1文のみが典型例に該当した）、調査する必要性が薄いと判断して、今回の調査対象からは除外した。

図2に「であって、」を含む典型例、図3に「において、」を含む典型例を示す。図2の例では、「であって、」の前の部分の名詞句を取り去っても、全体の名詞句の構造に影響を与えない。同じことは図3の例についてもいえる。

また、図2の例では、機能語の前の名詞句「ゴルフクラブ用グリップ」が、最後の名詞句と一字一句たがわ

ずに同じである。これを典型例としてある。図3の例では、機能語の後に改行があり、解析がしやすくなっているが、これについては、(e)で述べる。

ゴルフクラブのシャフトが挿入される断面概略円形の挿入孔を有するゴルフクラブ用グリップであって、前記挿入孔の内面に、シャフト挿入状態で押し潰されてグリップ外面部に軸方向に沿う凸筋が現出する細長い棒状の凸部を設けたことを特徴とするゴルフクラブ用グリップ。

図2 「であって、」を含む典型例（特開2003-180892）

遊技動作制御を統括する主制御手段と、この主制御手段が管理する遊技動作制御に必要な遊技データを格納する遊技データ格納手段と、遊技球の発射操作を行うための操作手段とを備え、この操作手段を遊技機本体と着脱自在に構成した遊技機において、前記操作手段側は、前記遊技データ格納手段が格納する遊技データを不揮発的に記憶するように構成されたことを特徴とする遊技機。

図3 「において、」を含む典型例（特開2003-180973）

表2 機能語の分類

		典型	非典型	その他	
				読点有	読点無
であって、	前回まで	112	17	9	2
	今回	29	4	9	1
	計	141	21	18	3
において、	前回まで	125	18	10	30
	今回	38	3	5	19
	計	163	21	15	49

遊技領域に設けられるものであって、上部に遊技球が流入する開口を備え、該開口の直下に遊技球検出器を具備する球検出路が連成された入球装置を備えるものにおいて、入球装置が、前記開口に、遊技球が所定検出位置を通過するように案内規制する球規制案内部材を設けたものであることを特徴とする入球装置を備えた弾球遊技機。

図4 「において」を含む非典型例（特開2003-180978）

表2に、機能語の分類例を示す。「前回まで」とあるのは、以前の調査450文についてまとめたものであり、「今回」とは、今回調査した100文の結果である。「典型」とは、図2、3に示したような、機能語の前の名詞句と、最後の名詞句とが完全に一致するような例であり、「非典型」とは、図4に示したように、機能語の前の名詞句と、請求項の最後に来る名詞句とが一致しないものである。また、「その他」とは、請求項の文の分離にわからない機能語の内訳を示す。この表からも分かるように、一般的には読点のつかない機能語は、比較的下の層の修飾句に関わっていることが多い。

### (b) 「特徴とする」の調査

「特徴とする」についても引き続き調査を行った。「特徴」という語は、今回調査した100文のうちに84文に出現し、「特徴する」（「特徴とする」の誤植？）と、「特徴とした」が各1文あった他は、すべて「特徴とする」という句として出現した。このうち多くは、図2、3のような典型例として、名詞句の最後の部分を修飾する形をとるが、機能語を伴わなかったり、「特徴とする」の後にさらに修飾句が入ったりする例もある。第4節で、機能語との関わりについて述べる。

導光棒の少なくとも一端面に第1発光体が配置され、前記導光棒内を透過する光を、その周面で屈折発光させる第1発光装置と、前記導光棒の背面側に離れて配置され、該導光棒に向けて発光する第2発光体が取付けられた第2発光装置と、から成ることを**特徴とする**遊技機の電飾装置。

図5 改行、「と」、「前記」などを含む例（特開2003-180950）

### (c) 助詞「と」による並立構造

名詞句の羅列ではなく、助詞「と」を介して、まとまった時点で全体が修飾句としてかかるといった例は請求項にはしばしば見られる現象である。図5に典型的な例を示す。この例では、2つの「と」が「成る」にかかる形をとっている。このような例は今回調査した100文の約6割で、これは前回の調査とほぼ同じである。

### (d) 動詞による並列構造

今回は比較的長い請求項が多かったせいか、動詞による並列構造と見られるものが、全体の8割近くをしめた。前回の報告で述べたように、「備える」や「有する」といった動詞も多いが、サ変動詞を含めて、あらゆる種類の動詞が用いられており、系統的に調査するのはなかなか難しいのが実情である。

### (e) 改行による係り受けの明確化

改行を入れることによって、係り受けがより明確になり、構文や意味解析が容易になるのは当然である。図5はその好例で、「発光装置」の並列が明確になっている。今回調査した100文でも、71文が文中に改行を入れており、そのいずれもが適切な改行であった。前回は6割近くが改行を入れていたので、だいたい6～7割がこれに相当すると考えられる。

### (f) 「前記」や「該」などによる階層構造

図5にも見られるように、「前記」、「該」、「当該」、「上記」といった照応的な名詞を用いて、句の中の係り受けやつながりを示す例は多い。前回は5～6割がそれに相当したが、今回調査した100文では、およそ8割にこのような記述が見られた。たとえば、「前記」という語を1文の中に複数用いている特許が75あった。ただし、これらの語は、句の中のいろいろな階層に出現するので、どのような役割を果たしているかを調査するにはなお研究が必要である。

## 4 機能語と「特徴」の組み合わせパターン

表3 機能語と「特徴」の組み合わせパターン

	であって		において		特徴 含まず
	典	非	典	非	
前回	36	9	59	8	35
今回	23	0	34	0	7
計	59	9	93	8	42

オペレータ自身の操作により運転自在な操作ハンドルを備えたパチンコ台の運転装置**において**、パチンコ玉の打ち出しを操作する操作ハンドルと、該操作ハンドルに設けられ、人体の接触を検知する人感センサと、該人感センサが人体の接触を検知したときのみ操作ハンドルからのパチンコ玉の打ち出し操作を実行するパチンコ玉打ち出し手段と、を具備してなるパチンコ台の運転装置。

図6 「特徴」を含まない例（特開2003-180972）

前回200の特許文について行った機能語と「特徴とする」の組み合わせパターンについて、今回も100文について調査した。その結果を表3に示す。

表3の「であって」の「典」とは、

・[名詞句A]であって、～を特徴とする[名詞句A]  
という形を示す。図2がその例である。「非」は前後の名詞句が一致しない場合を示している。

同様に、「において」の「典」も、

・[名詞句A]において、～を特徴とする[名詞句A]  
という形である。図3がその例となっている。「非」も「であって」の場合と同様である。

前回の統計では、「特徴を有する」や「特長とする」なども含めていたが、紛らわしいので、今回ははずしてある。前回の部分の数字が文献<sup>[1]</sup>と多少異なるのはそのためである。また、すでに上で述べたように、今回調査した中からは、「特徴する」や「特徴とした」も除外



してある。

表 3 から分かるように、対象 200 文を 100 文増加して 300 文に拡張してもほぼ同様の結果、すなわち、請求項の半数がこのようなパターンを持っていることが裏付けられた。

「特徴とする」を含まない例（表 3 のいちばん右側の欄）では、機能語によって分離する場合と、機能語があってもそこで分離しない場合、また、そもそも機能語もない場合とがある。機能語を含んでいて、その部分で分離するが、「特徴とする」が最後の名詞句の前でない例を図 6 に示す。一方、機能語がなく、「特徴とする」が含まれる例はこの表には含まれないが、図 5 に示してある。

## 5 問題点と今後の検討

人手で特許文を調査するのは、前時代的なやり方であるとして、最近はあまり行われないうえ、継続的にデータを観察していると、見えてくるものもあると考えられる。

また、ファインチューニングにこれらの知見をうまく生かせれば、もう少し特許文の請求項に即した解析を行うことができる可能性がある。

さらに、データがある程度たまってくれば、おおざっぱに分けたデータをグループ化して、順番に学習するといったことも考えられる。その場合、最も可能性のあるデータ群を最後に学習するのがよいとされているが、どうであろうか。

いろいろなパラメータの関連性や組み合わせについても今後さらに検討する予定である。

## 参考文献

- [1] 横山晶一：特許文請求項からの定型パターンの抽出と調査、2018 年度 AAMT/Japio 特許翻訳研究会報告書 (2019) pp.25-32,  
[http://aamtjapio.com/kenkyu/files/kenkyu05/AAMT\\_Japio\\_20190422.pdf](http://aamtjapio.com/kenkyu/files/kenkyu05/AAMT_Japio_20190422.pdf)
- [2] 須藤克仁：ニューラル機械翻訳の進展—系列変換モデルの進化とその応用—、人工知能 Vol. 34, No.4 (2019) pp.437-445
- [3] Alberto Poncelas, Gideon Maillette de Buy Wenniger and Andy Way: Transductive Data- Selection Algorithms for Fine-Tuning Neural Machine Translation, PSMT2019 (The 8th Workshop on Patent and Scientific Literature Translation) (2019)

