

# 接続情報を用いた名詞周辺の語の分類\*

丸尾聡子† 乾 裕子† 山岡美保 荻野孝野‡

†計量計画研究所 ‡日本電子化辞書研究所

## 1. はじめに

名詞は、品詞分類の上で「格に立つ語」という明確な基準があるため、副詞に比べて問題視されることが少ない。しかし、実際には名詞と他の品詞との境界は曖昧である。例えば「普通」という語は、辞書によって1) 名詞2) 名詞・副詞・形容動詞のいずれかと記述が異なっている。

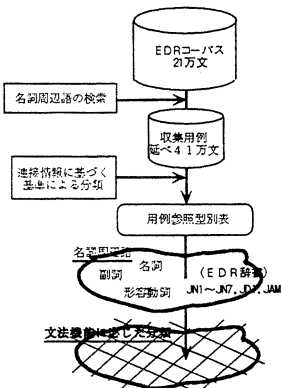
また、言語処理の立場から見ると、名詞周辺の語は形態素解析の際に適切に解析されないことがある。これは、名詞が副詞的に働く場合の十分な接続情報が解析用辞書に記述されていないためである。

これらの問題を解決するために、本研究では表層情報に基づき、実例分析によって名詞周辺語を細分化する。具体的には、1) 後接する助詞の種類、2) 副詞的用法の有無、また、副詞的用法を持つ場合には3) 後接する助詞の種類、4) 前接する語の種類に着目する。接続情報に基づく基準を立てることにより、語の役割に応じた区分が可能である。

本研究ではEDR日本語単語辞書(以下、EDR辞書)を対象とし名詞周辺語の網羅的な見直しを進めている。本稿では、副詞かつ名詞の役割を持つ語(以下、副詞的名詞という)の分析結果を中心に報告する。作業の流れは図1の通りである。

## 2. 関連研究

水谷[1]の研究では名詞周辺の語に関して、EDR辞書の品詞区分[3]に比べて細かい分類基準が設けられている。この研究は、接続情報により名詞・副詞・形容動詞を分類したものである。ここで簡単に水谷の研究について説明する。



目的: 連体詞・接続詞を

除く無活用

の自立

語を包括的な基準

で分類する。

方法: 各語のコーパス中の用法(接続情報)が下記の基準に適合すればフラグを立てる。そのパターンによりコードを付与する。

基準: 全部で15の基準がある。

結果: 延べ約10,000の実例を用い、約1,400語(連語)を分類。1,414見出し語が、93のパターンに分類された。

図1 接続情報を基準にした語の分類

種別: 情態類系統、純粋な副詞、ノヤナを許す副詞、相体言中核、名詞中核、副詞法のある名詞、名詞・相体言両用、など

表1 水谷研究での基準

基準	説明	語例
a	主格のガ、客格のヲ、相手・場所のニ、基準のカラの格に立つ	勢い、対応
e	格には立たないが、ノは続く。述語の要素になりうる。	必死、びったり
b	いわゆる形容動詞	明らか、裏腹
ノ	a,e以外でノがつく	正当、案外
ナ	a,eであり、連体修飾にも使える。	愚か、好都合
態	情態を示す語。1: 助詞がつかない。2: トを伴う	1: 大い、2: 堂々
ニ	名詞を副詞的連用修飾語に変えるニがつく	代り、夢中
ト	名詞を副詞的連用修飾語に変えるトがつく	不思議、自然
他	他のものがついて副詞的連用修飾語になる	夢え、お陰で
e5	独立して副詞用法がある。	結構、一方
c	連体修飾語を受けると、副助詞を伴わない副詞用法がある。	結果、時
c'	連体修飾語を受けると、副助詞を伴わない副詞用法がある。	以上、余り
陳	陳述の表現がある	万一、夢え
d	スルなどの語がついて動詞になれる	我慢、安心
転	体言・用言の範疇がなく、述語の要素となることが出来る語	尤も、随分

\*いわゆる形容動詞: 次の三基準を満たすもの。(品詞認識より)

- (1) 「一に」の形が無く動詞に対して副詞と同様の連用修飾の働きをすること。
- (2) 連体修飾語となる時の形が「一な」であること。
- (3) 「だろ・だっ・で・に・だ・な・なら」の活用形が原則として揃っていること。

この研究は、A) 従来の品詞に捕らわれずに接続情報から語を分類した点、B) これまで特筆されなかった基準(c,c',e,転)に注目している点、C) 実例を用いて分類している点で有用である。しかし、a) 内省による用法の補完を行っている点、b) 接続要素とコードが具体的に対応していない点に問題がある。本研究では、水谷の枠組みを適用して名詞周辺語を細分化する際、この二点の問題を解決する。

## 3. 調査方法

### 3.1 調査対象

EDR辞書の名詞周辺語を整理したところ、名詞・副詞・形容動詞に付与されていた品詞コードは次の通りである。

① JN1 かつ JN1 ② JN1 かつ JN4 ③ JN7 ④ JN4 ⑤ JAM  
これらの品詞が付与された語を調査対象とする。以下、各番号(丸囲み数字)で品詞コードを示す。なお、EDR辞書では表2のように品詞コードが付与されている(③より抜粋)。また各品詞の語数は表3の通りである。

表2に示す通り、EDR辞書では、同概念で名詞・副詞の両方の用法を持つ語、すなわち、いわゆる副詞的名詞には

\* 本研究は、EDR日本語単語辞書の保守補綴を発展させて行ったものである。

①から⑤の合計は 9.911 語である。

表2 EDR での品詞付与基準

品詞 コード	品詞	説明	例
JN1	普通 名詞	榴槤詞「が、を」が従受するもの	太陽、山
JN4	時詞	同じ概念で、名詞にも副詞にもなるもの。	今日、皆
JN7	形式 名詞	必ず連体修飾句に修飾されて成り立つ名詞。	こと、もの
JD1	普通 副詞	基本的に、そのままで、連用修飾語となること ができる。	すっかり、 ほほ
JAM	形容 動詞	終止形から終る活用語。連体形は基本的に ナであるが、ノを介して、あるいは直接ナ連 体句になる場合もある。	同じ(が)、 感動的(だ)

### 3.2 調査方法

EDR コーパス全文 (21 万文) から、3.1 で挙げた調査対象の語が出現する用例を全て収集する。こうして得られた用例を水谷の枠組みに従って人手で分類しフラグを立てる。これにより判断基準となる前接・後接要素の情報を一元的に扱うことが可能である。

対象とする語が形容動詞と考えられる場合には、水谷の挙げた 15 ケの基準に加え、名詞を直接連体修飾する用法の有無を調査する。後者の用法判別基準は EDR 辞書で採用されている。

分類の際には、実例から確実に抽出できる用法だけにフラグを立てる。実例に現れていなくても用法を想定できる場合があるが備考として記述するにとどめる。内省による補完は一切行わない。

ここで得られた結果をフラグの排列に従ってソートすることで、品詞の細分類を行う。

#### 4. 調查結果

#### 4.1 コーパスから用例が得られた語

対象とした9,911語を検索した結果、3,134語に関して用例が収集できた。詳細は表3の通りである。

表3 調査対象語数と用例が得られた語数

品詞 番号	対象の品詞	語数	用例が得られた 語数	割合	延べ例文数
①	JN1+JN1	128	44	34.38%	5,802
②	JN1+JN4	551	194	35.21%	20,534
③	JN7	103	86	83.50%	95,254
④	JN4	1,181	434	36.73%	29,862
⑤	JAM	7,948	2,376	29.89%	266,216
	合計	9,911	3,134	31.62%	417,664

用例が出現しない理由は、辞書見出しの表記と実例表記の差に負うところが多い。用例が収集できなかった語の9割近くは、下記 1)~4)のいずれかの理由による。

- 1) 見出し語が古い表記である（「此つ切り」「午後」）
- 2) 送り仮名のゆれ（「八分通り」「八分通」）
- 3) 辞書の見出し語が一般的でない（「極目」）
- 4) 見出し語の区切りが、コーパスでの区切りと対応しない（「第二次世界大戦末期」を見出し語としている）

## 4.2 用法の抽出

接続情報を判断基準にした用法の有無から用例参照型別表(表4)を作成した。[1]のコード付与基準に従うと、下記の例では最右列のようにコードを付与することが可能である。このコードは従来の品詞区分とは異なる。

表4 用例参照型別表 (例：副詞的用法のある名詞)

品詞 番号	見出し語	a	e	b	直ノ	ナ態	ニト	他	ε	5	c	c'	陳d	転	コード
①	各々	+							+						1001
②	みんな	+							+						1001
③	折	+									+	+			1006
④	前	+									+	+			1006
②	結果	+							+	+	+				1007
③	うち	+							+	+	+				1007
④	かわから	+							+	+	+				1007

内省的判断による品詞付与(表4 品詞番号)と、接続情報に基づく用法の分類(コード)の相違が明らかである。

今回、延べ758語、151,448文を調査した結果、EDR辞書では5品詞①～⑤の語群が36ヶのコードに分類できた。一般に副詞的名詞とされる語の特徴を分析するため、このうちJN4④を分類した結果を次節で述べる。

5. 品詞分類が十分でない語

### 5.1 副詞的名詞

### 1) 出現した用法

JN4 (副詞的名詞) の用法は、22 ケのコードに分類できた。ここから用例が少ないため必要な接続情報が抽出できないと作業者が判断したコードを除く<sup>1)</sup>。以上の作業によって 11 ケのコードを取り出した (表 5)<sup>2)</sup>。分類された語数が多いコード 1000, 1001, 1005 以外の語は、コードに出現した語を全て記述した。この作業でコードは用法の組み合わせの種類を指すので、以下、コードの異なりを用法ボタンと呼ぶ。表 5 に示す通り、従来、一つにまとめられていた語群を機能で分類すると複数のボタンがある。

## 2) 記述すべき連絡情報

副詞的名詞(JN4)と、名詞かつ副詞である語(JN1 かつ JD1)は、従来の国語辞典では同一辞書内で下記二通りに記される。

<sup>\*1</sup> が格に立つ用例がまたま出現しなかったために作業基準[a]と出来ず、基準[ノ]に分類した語(例:十日)の類が属するコードを削除した。削除したコードを付与した語の85%は収集できた用例が20例以下であった。

\*2 ただし、コーパスを変えるなどして、さらに用例を集めた場合、また、内省によって連発情報を助けた場合には、ここで分類した結果が変わる可能性がある。

表5 用例から抽出した用法パターンによる分類

コード	基準	用法パターン	語例
1000	a	副詞的用法なし	一角、一國、一着、街並み、初年、町並み
1001	a, ε5	名詞。単独で副詞的用法がある。	みんな、一瞬、一人ひとり、一生、一歩、一目、皆、各自、各人、近年、空路、月末、後日、史上、従来、瞬時、常日ごろ、先年、大昔、昼日なか、長期、長期間、長年、通常、冬季、内実、日常、白昼、平素、方々、ほとんど
1002	a, c'	名詞。連体詞に修飾されると副詞的用法がある。	気持ち、時刻、宵、明け方
1004	a, c''	名詞。連体修飾句に修飾されると副助詞を伴わない副詞的用法がある。	歳月、途上、傍ら、夕刻
1005	a, ε5, c''	名詞。単独で、また、連体修飾句に修飾されると副助詞を伴わない副詞的用法がある。	以後、以上、以前、一夜、現在、今ごろ、今後、今日、今年、昨今、昨秋、生涯、大正時代、当今、当初、日中、毎日、夜間、夜半、来月
1006	a, c', c''	名詞。連体詞、また、連体修飾句に修飾されると副助詞を伴わない副詞的用法がある。	時期、前
1007	a, ε5, c', c''	名詞。連体詞、また、連体修飾句に修飾されると副助詞を伴わない副詞的用法がある。	かたわら、夏、秋、早朝、途中、冬、当日、夜、夕方
1010	a=	名詞。名詞を副詞的連用修飾語に変える二がつく副詞的用法がある。	どさくさ
1011	a, 二, ε5	名詞。単独で、また、名詞を副詞的連用修飾語に変える二がつく副詞的用法がある。	元通り、一度
0141	ノ, 他, ε5	格に立たない。単独で、また、名詞を副詞的連用修飾語に変える二がつく副詞的用法がある。	单身
0000	陳 転	格に立たない。副詞的用法なし。	いくばく、千万

(i) 名詞（副詞的用法もある）

(ii) 《名・副》

この記述は冗長である。なぜなら、(i)(ii)の区別は、構文的ふるまいの違いを反映しておらず、表5で示した用法パターンを書き分けている訳ではないからである。また、(i)(ii)で区別された語群が同じ用法パターンを持つことがある。例えば「ほとんど」は(ii)と記されることが多いが、用法パターンは通常、辞書で(i)と記述される他の副詞的名詞と同じである（表5 コード1001）。特に差異は認められず、実例を見ても特殊な例はない。

出現する際の特徴に違いがなければ、計算機用辞書では、副詞的名詞に上記(i)(ii)のような区別は必要ない。むしろ

副詞的用法を持つ場合の表層情報を統一的に記述することが重要である。副詞的用法を持つ語を解析する際、下記(1)～(3)の情報を利用することが望ましい。

#### (1) 後接助詞

ある語が副詞的用法を持つ場合、1)助詞を伴わない、2)二を伴う、3)トを伴う、4)他の助詞を伴う、のように語によって後接助詞が分かれる。この違いを記述することで、どの後接助詞が副詞的用法かを区別することができるため構文解析の曖昧さを減らすことが可能である。コード1011に分類された「一度」を例に挙げる。

#1 一度に言われても困る。（副詞的）

#2 一度と言われても困る。

#### (2) 被連体修飾情報

副詞的用法を持つ場合に、どのように連体修飾を受けているかを明確に記述することで、下記の例も解析可能になる。例えば、コード1002に分類された「時刻」は、連体詞の修飾を受けた時は、助詞を伴わずに副詞的に働く。しかし、句レベルの修飾を受けた時には副詞的に働かない。この情報が記述されていれば、下記の例は、副詞的用法でなく名詞の並列であると解析することができ、曖昧さを減らすことが出来る。

#3 ～工程で、かつ異なった時刻、作業者によって、機械部品を製作し、任意に組み合わせで作るようになる。

#### (3) 名詞との連接

今回は行っていないが、名詞に後接するかどうか調査する必要がある。名詞と連接した時だけ助詞を伴わずに副詞的用法を持つ名詞がある。その情報を付与しておくことで、下記の例も適切に解析できる。

#4 撮影前、舞台稽古を見に行つて、あんまりすごいので驚きました。

上記をまとめると副詞的用法を持つ場合に記述すべき連接情報は以下になる。数値によるコード化も可能である。

#### 連接情報：

1. 後接要素

0:なし、1:二、2:ト、3:二・ト、4:他（具体的に記述）

2. 前接要素

0:なし、1:連体詞、2:連体修飾句、4:名詞、

n:このうちの複数（番号の和で示す）

表4 用例参照型別表からは、これらの情報を抽出でき、構文的ふるまいを反映した副詞的名詞の細分類として発展させることが可能である。

#### 3) 事例に基づくことの重要性

(1) 普通名詞とされる語

時・数量を表す名詞など、単独で副詞的に用いられることのある名詞の存在は知られている。しかし、時を表す名詞すべてに副詞的用法がある訳ではない。また、内省だけ

では副詞的用法を持つ名詞を見落とす可能性がある。よって、実例にあたることが重要になる。

例えば、EDR 辞書では JN1 だけが付与されている語のうち、「ふだん」「深夜」「少量」などは副詞的に使用されている用例が出現する。<sup>33</sup>

#5 ふだん、ニュートン内ではあまり見かけない世代でもある。

#6 深夜、マツムシが鳴く。

#7 ジルコニウム鉱石に少量含まれている元素で、中性子を吸収する能力が高い。

#7 で、「少量含まれる」は「少量(ガ) 含まれる」のガ格の無形化であると解釈し「少量」は普通名詞であると考えられることも可能である。しかし、名詞と動詞の接続を許す規則を作ることにより曖昧さが生じることは容易に想像できる<sup>34</sup>。従って、これらの語は副詞的名詞として区別し、実例から抽出した語ごとの接続情報を記述する必要がある。

## (2) 形式名詞

形式名詞が副詞的に働くことも多い。連体修飾を受けることから、形式名詞は確実に名詞である。しかし、一部の形式名詞は、連体修飾を受けることによって副詞的用法を持つという情報を記述する必要がある。

#8 そのつど、集会のテーマに合わせた脚本をつくってきた。

## (3) 名詞に後接する副詞

「以来」は格に立たないので名詞ではないが、純粋な副詞とも、ふるまいがやや異なる。単独、または名詞に後接した場合に副詞的用法を持つ語である。現在の EDR 辞書の接続規則では、名詞と副詞は接続できず下記の例は解析できない。

#9 同教育大の開校以来ずっと女子サッカーの指導をしている。

#10 5月以来の違法越境、そして今度のハンガリー政府公認の出国によって、東独政府は完全にメンツを失った。

名詞に後接して働く副詞があることから、一部の副詞には、名詞に後接して副詞的用法を持つという情報を記述することが望ましい。

以上のように実例にあたることで、見落とされている情報があることを指摘した。辞書の改良により解析の精度を上げるためには実例から十分な接続情報を抽出し、これらを細かく記述することが必要である。

<sup>33</sup> さらに「午前」は一般的には名詞に後接した場合(#1-1)以外では、副詞的用法を持たないと考えられているが、実際には、独立して用いられる用例(#1-2)も見られる。

#1-1 中止することを同日午前決めた。

#1-2 同日の交渉で午前、米国の農業補助金、輸入障壁とも原則的に全面撤廃する、との具体案を行っている。

<sup>34</sup> 一般的には「少量、含まれる」のように表記されるため接続可能になっていることが多い。しかし、実際はコーパスにあたる読点を介さず表記する例が頻出する。これらも解析できるようにするためには、語の品詞・接続属性を十分に検討する必要がある。

## 5.2 その他

名詞周辺の語を整理するにあたっては、副詞的名詞だけでなく、以下のような問題がある。

・副詞とサ変動詞の扱い(例:さっぱりする、いきいきする)

・副詞的に働く形容動詞(例:同様)

・対応する語の品詞付与の整合性

(例:春・夏・秋・冬に異なる品詞が付与されている。

用法の分布は同じ。)

このように検討すべき課題は多い。接続情報付与の観点からは下記の二点を挙げる。これらの語は用法パターンに特徴があり、接続情報を組み入れることで解析の精度は上がると考える。

### 1) 格に立たない名詞

[1]では、ガ格・ヲ格に立せず、かつ形容動詞語幹でもないものをe型とした。今回の調査でも多く出現したが、これらの語に付与されている品詞・接続情報は他の普通名詞と変わらない。

例:絶好、同時、高級、いろいろ、いっそう、いっぽう、それなり、久々

これらの語は、ガ・ヲが接続しないため純粋な名詞と言えず、接続情報は他の普通名詞と変えることが望ましい。これらを名詞として扱うのであれば、格に立たない名詞も存在することが分かる。

### 2) 連体修飾する際の助詞(後接可能な助詞の制限)

名詞が連体修飾をする場合のナ・ノは、使い分けがある。一般にはノで連体修飾することが多いが、語によってはナがこの機能を持つ。これらを記述することで語の情報はより正確になり、解析時には有効である。

## 後接要素

1:ノ、2:ナ、3:ノ・ナ

#11 前に出るのが苦手な、内気な少年だった。

#12 \*前に出るのが苦手な、内気少年だった。

## 6. おわりに

以上、コーパスに現れた用例を接続情報で分類した結果に基づいて、副詞的名詞の構文的なふるまいについてまとめた。また、名詞周辺の語を実例をベースにして分類することの重要性を述べた。今後は、5.2で挙げたその他の品詞について、同様の比較検討を行う予定である。

謝辞: コーパスからのデータ収集に協力して下さった東京工業大学大学院情報理工学研究所の白井静昭さんに感謝致します。

## 参考文献:

[1]水谷静夫・星野和子(1994)「名詞から副詞まで一語類の新しい枠づけ」『言語学』, vol.19 no.7

[2]水谷静夫(1995)『包括的基盤による名詞・副詞類の仕訳』IPAシンポジウム講演会資料

[3](株)日本電子化辞書研究所(1993)『EDR 電子化辞書使用法調査』

[4]水谷静夫(1991)『稿本文法大体系』東京女子大学日本文学科

[5]西尾実他編(1994)『岩波国語辞典第5版』岩波書店

[6]萩野紫穂(1988)『標準語の文法』東京女子大学日本文学科