

対訳コーパスにおける品詞タグ
～名詞属性を持つ日本語品詞～
塚脇 幸代
s-tsuka@dream.ocn.ne.jp

0. はじめに

日本語と英語の対訳コーパスに、言語間の対応を観察する手がかりとして品詞タグを付与する際、各言語で伝統的に用いられてきた品詞体系を採用すると、相手言語には存在しない、その言語に固有の品詞が現れることになる。無論、各言語の品詞体系は、それぞれの言語のために考案されたものであって、対となる言語を想定すべき義務は無い。機能語においては、該当言語が表層にどのような機能的語彙を要求するかにより、どちらかに必要でどちらかには無用であることがある。日本語における助詞類、英語における前置詞、冠詞等は、文における明確な役割を持ち、それぞれの言語において必要な要素である。

日本語と英語の対訳においては、内容語にも、日本語側に特徴的な品詞が存在する。本稿では、そのうち、形容動詞と呼ばれる範疇と、それに隣接した関係にある副詞を取り上げる。¹

1. 形容動詞

典型的な形容動詞の特徴として、以下の3つの用法を持つ。

1) 述語用法がある

～は(形容動詞語幹)だ: 静かだ

2) 限定(連体修飾)用法がある

(形容動詞語幹)な(なる)～: 静かな

3) 副詞用法がある(連用形)

(形容動詞語幹) + に/と : 静かな、堂々と

ただし、これだけでは説明のつかない形容動詞語幹も存在する。「緊急」は、「緊急な」「緊急の」という2つの、連体修飾形をもつ。さらに、「緊急を要する」「緊急性」のように、助詞「を」、接辞「性」が後接可能という、名詞の特徴を顕わす。

ChaSen²はこれを「名詞-形容動詞語幹」と解析している。述語用法「～だ」や、連体用法「～な」が存在し、かつ、名詞のような語構成をとることを考慮した分類といえる。

しかし、文構造を記述する際にはもう少し統合された要素単位とラベルが必要になる。さらに、コーパスというこの形態素が現れる環境においては、それが名詞的用法なのか、副詞的用法なのか、一意に決定できる条件は整っているはずである。対訳であれば、他の言語との対応がとりやすい形にしておきたいという要求もある。

第一次解析に利用した ChaSen は、「大いに」などの典型的な形容動詞については、連用形を副詞として解析結果を出力する。その他に関しては、個別の形態素として出力する。

対訳コーパスで使用する品詞タグとしては、可能な限り名詞、形容詞、副詞等、相手言語である英語にも存在する品詞として扱うようにした。その例を表1に示す。なお、本稿で扱う品詞タグ体系では、名詞性と形容詞性を併せ持つという意味で、形容動詞語幹に AN というラベルを用いている。³

¹本文中の例の一部は独立行政法人情報通信研究機構自然言語グループの「日英新聞記事対応付けデータ」から取り上げた。

²日本語形態素解析システム 茶筌(ChaSen) version 2.1 for Windows, ©奈良先端科学技術大学院大学自然言語処理学講座

³品詞体系の詳細については参考文献1を参照のこと。

パターン	組み合わせ	対訳コーパスにおける扱い	例	ChaSen の解析例 (形容動詞語幹に対して)
Z	形容動詞語幹+だ	AN+だ	「簡単だ」	名詞-形容動詞語幹
A	形容動詞語幹+な	ADJ	「単純な」	名詞-形容動詞語幹
B	形容動詞語幹+に	ADV	「純粋に」	名詞-形容動詞語幹
C	形容動詞語幹	ADV	「当然、太郎は来る。」	副詞-助詞類接続
D	形容動詞語幹+の	RT	「突然の訪問」	副詞-助詞類接続
E	形容動詞語幹+する	VN	「ゆっくりする」	副詞-助詞類接続
F	形容動詞語幹+格助詞	N+JK	「安全を確認する」	名詞-形容動詞語幹

表 1

これらのパターンをすべて持っている形容動詞は稀である。その例をいくつか表 2 に挙げる。

	Z	A	B	C	D	E	F
寛大	寛大だ	寛大な	寛大に				寛大を旨とする
ゆっくり	ゆっくりだ	ゆっくりな	ゆっくりと	ゆっくり		ゆっくりする	ゆっくりがいい
高度	高度だ	高度な	高度に		高度の		
スムーズ	スムーズだ	スムーズな	スムーズに				
多様	多様だ	多様な	多様に		多様の		
平和	平和だ	平和な	平和に				平和を望む
完べき	完べきだ	完べきな	完べきに				完べきを求める
危険	危険だ	危険な					危険を冒す
緊密	緊密だ	緊密な	緊密に				
グローバル	グローバルだ	グローバルな	グローバルに				
顕著	顕著だ	顕著な	顕著に				
厳密	厳密だ	厳密な	厳密に				
最後	最後だ		最後に	最後	最後の一葉		最後を飾る
さまざま	さまざまだ	さまざまな	さまざまに	さまざま			
迅速	迅速だ	迅速な	迅速に				
すぐ	すぐだ		すぐに	すぐ	すぐの発車		すぐがいい
大事	大事だ	大事な	大事に				大事をとる
確か	確かだ	確かな	確かに				
適切	適切だ	適切な	適切に				
不安	不安だ	不安な					不安を抱く

表 2

表 2 において、Z は述語性 (V 属性)、A, B, C, D が修飾性 (A 属性)、E, F は名詞性 (N 属性) である。

ChaSen の解析では副詞になっているものもあるがここではこれらも形容動詞語幹 (AN) と分類する。その根拠は、述語性である。「～だ」で状態を表せるかどうかで判断する。たとえば、「トマトは野菜だ。」と「太郎は完べきだ。」とでは、前者は "A is N." の関係だが、後者は "A is ADJ." の関係で表される。これは意味の範疇になるのであろうが、AN を付与することにより、述部の性質を示すことができる。

表中には、人によって適合性に疑問をもたれる表現もあると察せられる。「ゆっくり」は本来トタル活用であるから、「ゆっくりな」という言い方は無いはずである。にもかかわらず、Web 上で検索すると決して少なくない件数がヒットする。また、「すぐの」という言い方には、筆者は違和感を覚えるが、表中に示した例は、駅の構内アナウンスで日常的に耳にするものである。このような事実は、「～な」や「～の」の造語力の強さを示すとともに、形容動詞語幹 (AN) の名詞的性質を裏付けるものと考えられる。

2. 副詞

副詞の条件として、

- 1) 活用が無い
- 2) 修飾先がある

が挙げられるが、その様相は多岐にわたる。

表3は副詞(句)がどのような形態素で構成されるかを示す。副詞はさまざまな形態素で構成され、一形態素で副詞のみの用法をもつものは限られている。「本日」、「現在」など、名詞としても使われる語彙は、前節の形容動詞のパターンに挙げたように、形容動詞の語幹がそのまま副詞となる場合もある。同時に、「に」が後接して副詞となる場合もある。この両方をもつものは名詞-副詞可能と解析され、形容動詞と同じく、二通りの品詞の可能性を持っている。

	RT	N	AN	VN	ADJ	ADV	V	AUX	JK	JF	JS	N-comp
どのように	どの	よう							に			
最も						最も						
あまりに		あまり							に			
そのためには	その	ため							に	は		
もっと						もっと						
強く					強い							
このような観点から	このよう な	観点							から			
すべて			すべて									
容赦なく				容赦				ない				
直接			直接									
現在			現在									
本日			本日									
より						より						
この目的のため	この	目的							の			ため
緊密に			緊密						に			
引き続き							引き 続く					
相互に			相互						に			
なぜ						なぜ						
ますます			ますます									
続いて							続く				て	

表3

また、副詞はそれと認定するのに、文構造の解析を待たなければならない場合がある。それは副詞の形態が多様であることに起因する。「太郎は続いて休んだ」と「雨が続いて不作になった」とでは、「続いて」の品詞を使い分ける必要がある。先に述べた「形容動詞語幹+に」についても、パターンBとパターンFの両方の可能性をもつものは、後接の「に」のみでその用法を特定することが難しくなる。

副詞の修飾先として、述部、文全体、名詞句があげられる。文全体を修飾する、いわゆる文副詞は、接続詞、接続助詞と接点を持つ。また、名詞を限定する副詞として、「ほとんどの学生が勤勉だったが、とりわけ太郎は熱心だった。」のような、取立ての機能は、係助詞や副助詞(JF)の機能と関わってくる。さらに、「今後三年間」などのように、数量表現に関与するものがある。

3. 品詞タグ体系における位置づけ

日本語の形容動詞については、その形態と用法が必ずしも一対一の関係にないといえる。そのため、

形態的に形容動詞であっても、文における役割は、名詞であったり、形容詞であったりする。副詞はその形態と用法とがさまざまであるため、他の品詞の可能性を排除した後、副詞と認定できる場合も多い。A(修飾要素), N(名詞), V(動詞)を基本とした品詞タグの体系に、形容動詞、副詞を位置づけると、以下の図1ようになる。

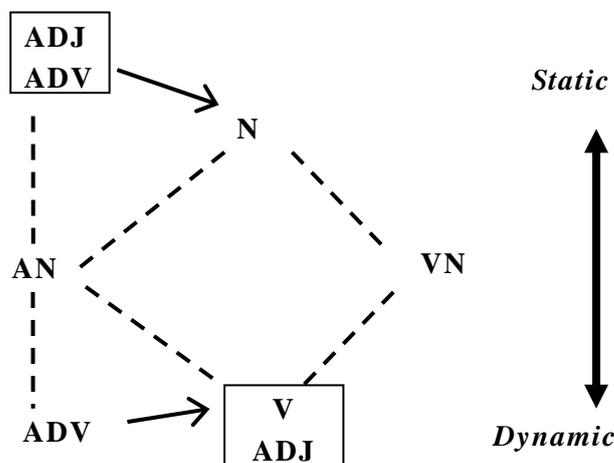


図1

図1の左部において、Nは名詞、Vは動詞、VNはサ変動詞語幹、ANは形容動詞語幹、ADJは形容詞、ADVは副詞である。破線は形態の遷移の可能性を、実線の矢印は修飾関係を表す。右部の両矢印は静的表現か動的表現かの極性を示す。Vと同じ位置にあるADJは形容詞、形容動詞の述語用法、ADJと同じ位置にあるADVは名詞を限定する場合の副詞であるが、ADVの機能の定義を動的表現に対する修飾とすると、別の名前で分類すべきかもしれない。そのためには、もっぱら名詞を修飾する連体詞との関係性を検討する必要がある。

4. 今後の課題

実際に運用された文や発話には、当然だが品詞はついていない。人間がその意味を理解するためには、発せられた形態や音声を手がかりにするしかない。それはそれぞれの言語によって異なり、また、それぞれの言語に固有の事情と規則がある。対訳を視野に入れる前に、個々の言語(日本語)の事情を整理している段階である。形態と文法的機能とは、必ずしも対応しないことを前提とした品詞体系の必要性を感じている。さらに複数の言語、すなわち当面は英語にも適用できる体系を、コーパスの分析によって進めていく予定である。今回取り上げなかった日本語の連体詞をはじめ、英語品詞、日英両言語の機能語については、別の機会に考察する。

<参考文献>

1. 塚脇幸代. 対訳コーパスのためのタグ体系, 言語処理学会第11回年次大会. P3-1. 2005年3月.
2. Tsukawaki, Sachiyo: Extracting Japanese-English Transfer Patterns from a Tagged Parallel Corpus. Graduate School of Language Education and Information Science, Ritsumeikan University. January, 2005.