

# 英文作成支援システムにおける 5 F - 8 冠詞の推敲 — — 名詞の分類法

小澤 邦昭 絹川 博之  
(日立製作所 システム開発研究所)

前田 多章 神成 広之  
(日立マイコン)

## 1. はじめに

先に、英文作成支援システムが具備すべき機能を提案した<sup>1)</sup>。その中の一つに、推敲支援機能がある。日本人の書く英文には冠詞の誤りが多いことはよく知られており、本報告では冠詞の推敲方法を取り上げる。冠詞の推敲は名詞の持つ性質と密接な関係があるので、ここでは、名詞の可算・不可算性をベースに、複数形の表現(語尾変化)、単複形と単複扱いの対応のねじれ、冠詞の使用の制約を基準にして、名詞の体系的分類法を論じる。

## 2. 冠詞の付け方の原則と名詞の分類の基本型

冠詞の付け方の原則は次の通りである(なお、無冠詞とは他の限定詞も伴わないことを、ここでは意味する)。

- (1) 単数可算名詞は、単数形を取る。不定冠詞を伴うことはできるが無冠詞で用いることはできない。
- (2) 複数可算名詞は、複数形を取る。無冠詞で用いることはできるが不定冠詞を伴うことはできない。
- (3) 不可算名詞は、単数形を取る。無冠詞で用いることはできるが、不定冠詞を伴うことはできない。

この原則に従って、名詞を、単数扱い、複数扱い、不可算扱いできるか否かを基準にして、分類すると、表1のように7つの基本型が得られる。この基本型により、上述の冠詞の付け方の原則に違反した名詞を見つけることができる。例えば、information 型の名詞が、複数形であったり、不定冠詞をとる場合には、上述の原則(3)に違反するので、推敲対象として検出できる。

## 3. 基本型の拡張の必要性

### 3.1 語尾変化の不規則現象

複数形は、原則として、単数形に -(e)s がついでいることで判定できる。しかし、sheep (time型)や city (book型)のように、複数形が不規則変化(無変化を含む)する例外語もある。逆に、複数形が不規則変化する場合には、単純に -(e)s を付けて複数形を作ると誤りとなる。このような例外語の誤りを検出するために、語尾変化の情報を基本型に付加する必要がある。

### 3.2 単複形と単複概念のねじれ現象

単数形は単数扱い(単数可算、不可算)に対応するのが原則であるが、集合名詞は単数形でありながら複数扱いできる。例えば、book型に属する集合名詞 groupは個々の構成要因を意味するときには、単数形のまま複数扱いされる。しかし、複数扱いされるにもかかわらず、無冠詞で使用することはできず、前述の原則(2)の例外となる。これは、単数形で複数概念を表わすときには、複数概念が十分に強くなっていないためと考えられる。複数扱いされるにもかかわらず、無冠詞で使用できないことを複数概念「弱」と定義し、原則通り無冠詞で使用できることを複数概念「強」と定義する。people も同じ book型の集合名詞であるが、複数概念「強」の例である。従って、単数形で複数概念の表現の可否、及び可のときの複数概念の強さの情報を基本型に付加する必要がある。

一方、複数形で単数扱いの可否を示す情報も必要である。この時、単数扱いとして、単数可算と不可算とを区別する必要がある。例えば、mean (手段)の複数形は単数可算として扱うことができ、domino (ドミノ遊び)の複数形は不可算として扱われる。この場合、単数可算にもかかわらず不定冠詞をとらない現象、あるいは不可算にもかかわらず無冠詞で用いることができない現象があれば、単数可算概念の強さ、不可算概念の強さを、それぞれ定義する必要が出てくる。

### 3.3 制約現象

名詞によっては、常に定冠詞あるいは不定冠詞をとるものがある。例えば、information型の名詞 equator は常に定冠詞をとり、feel型の名詞 standstill は常に不定冠詞をとる。また、police は通例、定冠詞をとる。このような制約情報を基本型に付加する必要がある。

表1 名詞の基本型

基本型	代表的名詞	単数扱い	複数扱い	不可算扱い
1	time	○	○	○
2	book	○	○	
3	knowledge	○		○
4	energy		○	○
5	feel	○		
6	police		○	
7	information			○

#### 4. 基本型の拡張 — 例外を考慮した名詞の分類

基本型には3つの属性（語尾変化、単複形と単複扱いの対応、制約）を与え、各属性値を次のように定義する。

[語尾変化]

- 0…規則変化（単数形に-（e）sが付く）
- 1…複数形をとらない名詞か、複数形のみ名詞で語尾変化に関係無し（feel, police, information等）
- 2…不規則変化（sheep⇒ sheep, city⇒ cities等）
- 3…不規則と規則変化の両方（index⇒ indices, indexes）

[単複形と単複扱いとの対応]

- 0…規則的対応（単数形⇒単数扱い、複数形⇒複数扱い）
- 1…単数形で複数扱いされる（複数概念は「強」）
- 2…単数形で複数扱いされる（複数概念は「弱」）
- 3…複数形で単数可算扱いされる
- 4…複数形で不可算扱いされる

[制約]

- 0…制約無し
- 1…常に定冠詞を伴う（equator, brink, Pacific等）
- 2…常に不定冠詞を伴う（modicum, standstill等）
- 3…常に無冠詞で用いる（Japan, God, Jupiter等）
- 4…通例、定冠詞を伴う（police, outskirts等）
- 5…通例、不定冠詞を伴う（think, liking等）
- 6…通例、無冠詞で用いる（research, television等）

制約の属性値2～3の「常に」という意味は、文脈によって定冠詞を取る場合を除外している。また、制約の属性値4～6については、これに違反しても必ずしも誤りとはいえないので、下記の表2の分類からは除いてある。

#### 5. 名詞の分類結果の検討

当社の機械翻訳システム HICATS/EJ (Hitachi Computer Aided Translation System / English to Japanese) の基本辞書（約5万語）の名詞を基本型に分類した結果、

book型が一番多く約1万2千語、次いで information型が約8千語、time型が約4千語、その他の型は合わせても約300語であった。属性値を考慮したときの組合せを単純に計算すると、名詞は560通り（基本形7×語尾変化4×単複の対応5×制約4）に分類される。しかし、実際にはこれらの組合せがすべて起こるわけではない。これまでに報告者らが見つけた組合せは24通りである（表2参照）。この組み合わせの数は、今後もそれほど大きく増加することはない見通しである。

この分類体系は、名詞一語の性質に着目しており、他の語との共起を考慮した情報は取り込んでいない。例えば、time型の名詞は、単数、複数、不可算のいずれの取扱いも可能であるので、冠詞の付け方の原則に基づいて誤りを見つけることはできない。この型の名詞は、共起情報の利用により効果的に推敲できる場合がある。例えば、熟語表現 kill time を利用すれば、kill a timeや kill the time 等の誤りを推敲できる。

#### 6. おわりに

冠詞は日本人が苦手とするにもかかわらず、冠詞推敲の従来研究は少ない<sup>2)</sup>。本報告では、冠詞の推敲の基本となる名詞の分類の体系化について述べた。この分類は、可算・不可算と複数形の語尾変化の観点から見た冠詞の付け方の原則に基づき、例外現象も取り込んでいることが特徴である。これにより、冠詞の明らかな付け誤りを検出し、かつ、正しいのに誤りとして検出する第2種の誤りを防ぐことができる。ただし、a, the, 無冠詞のいずれもが文法的に許容される場合に、適切な冠詞の決定を支援することは、今後に残された課題である。

#### 参考文献

- 1) 小澤邦昭ほか：情報学会第38回全国大会予備集、PP. 220-221(1987)
- 2) 河合敦夫ほか：情報学会論文誌、vol.25, No.6, pp. 1072-1079(1984)

表 2 名 詞 の

番号	基本型	語尾変化	単複の対応	制約	代表的名詞
1	1	0	0	0	time
2	1	0	1	0	youth
3	1	0	2	0	family
4	1	0	3	0	work
5	1	0	4	0	domino
6	1	2	0	0	sheep
7	1	3	0	0	focus
8	2	0	0	0	book
9	2	0	1	0	people
10	2	0	2	0	group
11	2	0	3	0	mean
12	2	2	0	0	city

分 類 結 果

番号	基本型	語尾変化	単複の対応	制約	代表的名詞
13	2	2	0	1	backwoods
14	2	3	0	0	index
15	3	1	0	0	knowledge
16	4	0	0	0	energy
17	4	2	0	0	statistes
18	5	1	0	0	feel
19	5	1	0	2	standstill
20	6	1	0	0	police
21	6	0	0	0	thank
22	7	1	0	0	information
23	7	1	0	1	equator
24	7	1	0	3	Japan