

EDR電子化辞書における英語複合語・慣用表現の記述方法

6F-6

石渡裕美、有岡昌子、原田千秋、天野真家

(株)日本電子化辞書研究所

1. はじめに

(株)日本電子化辞書研究所(EDR)では、自然言語処理用の大規模電子化辞書の開発を行っている。本稿では、このうちの英語単語辞書における複合語・慣用表現の記述方法について述べる。

ここでいう複合語・慣用表現とは2語以上からなる特定の言い回しであり、連語も含む(以下複合語と略す)。活用、挿入等の変形、特定の語や形態素に特定できない変数部分の記述などに対応するため、複合語は字句通りの一語登録ではなく、構成語の組合せとして記述する必要がある。この時、個々の構成語の統語関係や制約情報などの記述の方法の一貫性を保つことが重要である。

2. 構成語の組合せとしての複合語の記述

(1)「見出し情報」として、複合語の構成語の並びを記述する。検索見出しは複合語中の最初の構成語の不変部分とする。

(2)「文法情報」として、複合語全体についての文法情報及び構成語間の構文的な関係を記述する。品詞は、複合語全体としての品詞を付与する(図1)。また、構文木によって構成語間の構文的な関係を示す。その他、次に示すような構文や語の属性に関わる情報を記述する。

名詞：動詞や代名詞と呼応する場合の数や性に関する一致、冠詞との共起、

動詞：文型、相、態、法

形容詞・副詞：他の構文要素との修飾関係、比較表現の有無、

3. 複合語記述の多様性への対応

複合語の記述において、複合語やその構成語のふるまいを表現する方法は多様である。多くの場合その中から一つを選択する絶対的理由は少ないのであるが、記述者によるばらつきを回避し、一貫性を保たなければならない。そのため、多様な記述方法があるケースについては記述方法を一つに決める必要がある。以下にその例を示す。

3.1 語形に制約のある構成語への対応

複合語中に「常に複数形で用いられる」などの語形制約のある構成語が含まれる場合(eg. 'hands down': 容易に)、このような構成語の記述の方法として次の3通りが考えられる。

①複数語尾's'までを含めた複数形の形を一構成語として記述→'hands'

②単数形で記述し「常に複数形で用いられる」との情報を付加→'hand' (常に複数形)

③単数形と's'をそれぞれ一構成要素として記述→'hand/s'

この複合語内でその語が「複数」であることは必須であるが、①では複数形を表層の文字列として表現するため、②では語の属性として表現するため、③に比していずれも多少不自然である。したがって、その語が常に複数形で用いられることは③の方法で示す。常に過去分詞形で用いられる動詞の語尾'ed'なども同様で、必須の語尾は特定の要素としてその存在を明記する。

'The Method of Describing English Compounds and Idioms in the EDR Electronic Dictionary'
Hiromi Ishiwatari, Masako Arioka,
Chiaki Harada, Shin-ya Amano,
Japan Electronic Dictionary Research Institute,
Ltd.

動詞 (eg. get into w#something : w#something
の中に入れる)

動詞句 (eg. kick the bucket : 死ぬ)

名詞 (eg. air conditioner : エアコン)

名詞句 (eg. cakes and ale : 人生の楽しみ)

形容詞句 (eg. wide awake : すっかり目覚めて)

副詞句 (eg. once upon a time : 昔々)

前置詞句 (eg. from time to time : 時々)

群前置詞 (eg. in front of : ~の前に)

図1

3. 2 変形を伴う複合語への対応

ある複合語が変形を伴うものであることは、見出し情報に示す。一つの単語や形態素に特定できない変数部分の記述には、品詞やその指示する内容を代表させる語を用いる。修飾語など複合語外の要素の挿入の可能性も構成語の並びの中に示す。

【例】make use of ~ : ~ を利用する

→ mak/ w#語尾/ //use/ /of/ /w#something

(検索見出し)

(構成語情報)

(‘w#~’は変数部分を示す。)

(‘use’の前の//は、ここに‘heavy’などの必須ではない要素の挿入の可能性を示す。)

3. 3 受動態表現への対応

受動態表現(‘be’+過去分詞)の記述には、

①能動態の形に戻し「常に受動態で用いられる」との情報を付加

②‘be’を含めた受動態表現の形を記述

③‘be’を除き、過去分詞で始まる形を記述して「常に受動態で用いられる」との情報を付加

の3通りの方法が考えられる。①を採ると、能動態と受動態では構成語が異なるため、記述した情報は変形しなければ用いることができない。②では‘get’など‘be’以外の動詞を用いた表現に対処できない。したがって見出しとしては‘be’を除いた③の形を設定する。

3. 4 冠詞情報への対応

例えば(the United States of America) (アメリカ合衆国)は、通例‘the’を伴って用いられるが、実際の言語運用上は‘the’を伴わず使用される場合もある。このような語用論的な問題に対処するため、見出しには冠詞を除いた形を登録し、「一般に‘the’を伴う」との情報を付加する。

3. 5 “-ing”形を持つ語への対応

“-ing”の形を持つ語には、

①動詞の活用形とみられるもの

②派生語とみる方が適切であるもの

(eg. ‘interesting’)

がある。それぞれ、①は“~(動詞)/ing”として、②は一語としてその働きに応じ形容詞または名詞として、記述した。

4. 記述作業上の問題

語のふるまいに関して、記述作業時に判断の揺れが生じることがある。特に複合語の場合は、構成語間の結合の強さや複合語としての固定化の度合いに関する判断の揺れの影響を受けやすい。

例えば、前置詞句の中には比較表現を持つものがあるが(eg. ‘on the ball’:有能な、適している→‘more on the ball’)、比較表現を持つか否かは記述者によって判断の揺れが生じやすい。比較表現があるような概念か否かの判断の難しさという点においてこの問題は単一語の形容詞・副詞の場合と同様であるが、ここでは更にその複合語が一つの表現としてどの程度固定化されているか、という問題が加わる。

また、構成語間に修飾語などの挿入が許される複合語と許されない複合語が存在する。これは複合語の結合の強さにより決まるものであるが、この判断に関しても判断の揺れは生じる。

5 まとめ

以上、EDRにおける英語複合語の記述方法について述べた。複合語記述の多様性への対応方法を明確にしながら記述作業を進めることにより、記述の一貫性を保つことができる。

【参考文献】

- [1]A.S.Hornby著、伊藤健三訳注；「英語の型と語法」オックスフォード大学出版局(1977)
- [2]末松博他；「英語複合語・慣用表現の辞書表現法」情報処理学会第43回全国大会講演論文集 3-217(1991)
- [3]日本電子化辞書研究所；「TR-019 英語単語辞書」日本電子化辞書研究所(1990)