

## 5B-5

情報処理用語における複合語の  
構成パターンの抽出とその利用について

・須之内美幸、石井敬子、熊野明、三輪和弘、塚田康博、小寺次夫  
(株)日本電子化辞書研究所

まえがき

本発表は、(株)日本電子化辞書研究所(以下「EDR」と略す)における専門用語辞書[1]の開発経過報告を行う。情報処理用語は複合語が多いことが特徴であり、複合語を構成語へ自動分割することは難しい。複合語の構成パターンの抽出を行い、抽出したパターンを利用して情報処理用語における日本語の複合語の概念記述データの作成方法について検討した。特に対訳情報中の前置詞について検討を行った。効率的な概念記述の作成に役立つと考えている。

## 1. 情報処理用語の特殊性とパターン抽出

情報処理用語の特徴を以下に示す。

- ・見出し語の90%が複合語である。
- ・情報処理用語において頻繁に使われる用語は限定されている。これを中心語とよぶ。
- ・中心語に周辺概念の用語を修飾し、多くの用語が作られている。  
中心語の例 ; 「program」, 「system」
- ・複合語は、2~4語から構成されることが多く、名詞の連鎖形が多い。
- ・英語の発音をカタカナ表現した日本語が多い。
- ・日本語の複合語の名詞は、英語の名詞句と1対1で対応することが多い。

日本語、英語の見出し語をパターン解析した結果を表1に示す。

日本語のパターン	英語のパターン
機械翻訳 【名詞+名詞】	machine translation 【名詞+名詞】
電子スイッチ 【名詞+名詞】	electronic switch 【形容詞+名詞】
制御器の割込み 【名詞+助詞+名詞】	interrupt of controller 【名詞+前置詞+名詞】

表1. 語の連鎖例

On using and analyzing for patterns of compound words in the information processing term  
Miyuki SUNOUCHI, Noriko ISHII, Akira KUMANO, Kazuhiro MIWA, Yasuhiro TUKADA, Tuguo KODERA  
Japan Electronic Dictionary Research Institute, Ltd.

## 2. 複合語構成パターンを利用した概念記述

概念記述作成作業には、語と語の共起関係のデータが基本となる。[2],[3] 表層情報を基に語と語の格関係を関係子を用いて記述し、共起関係を作成する。複合語が名詞の連鎖の場合、構成語への分割が困難である。日本語の共起関係を機械的に抽出する場合、構成語への分割が複数考えられる。構成語に分割しても、名詞と名詞の間に殆ど総ての助詞を挿入できる。助詞に応じて複数の意味が成り立ってしまい、関係子を特定できない。

情報処理用語の特殊性で述べたごとく、情報処理用語の複合語が名詞で構成されている場合、その対訳と1対1で対応することが多い。また、情報処理用語の複合語は、その中心語をもとに構成語への分割ができる場合が多々ある。つまり、上記問題点に対して、情報処理用語の場合、複合語の構成語への分割が比較的可能である。さらに情報処理用語は、英語で「名詞」+「前置詞」+「名詞」のパターンを取る複合語が日本語に翻訳されるとき前置詞が訳されない場合が多い。

構成語の格関係を補うために、訳語中の前置詞を利用した、新しい複合語の分割手法を提案する。

## 3. 実験結果

複合語を分割し、関係子を付与した作業手順を以下に示す。

- ステップ1. 情報処理用語の日本語の複合語を中心語をもとに構成語に分割する。
- ステップ2. 対応する情報処理用語の対訳情報の前置詞を参照する。
- ステップ3. 前置詞を格助詞に置き換える。
- ステップ4. 表層格と助詞の対応テーブルをもとに、関係子を選ぶ。

‘テキスト終結’、‘位置参照’及び‘内容検索’について行った各ステップの結果を表2に示す。

ステップ	テキスト終結	位置参照	内容検索
1	テキスト/終結	位置/参照	内容/検索
2	end of text	reference by location	search on context
3	テキストの終結	位置 による参照	内容の検索
4	終点	方法	限定

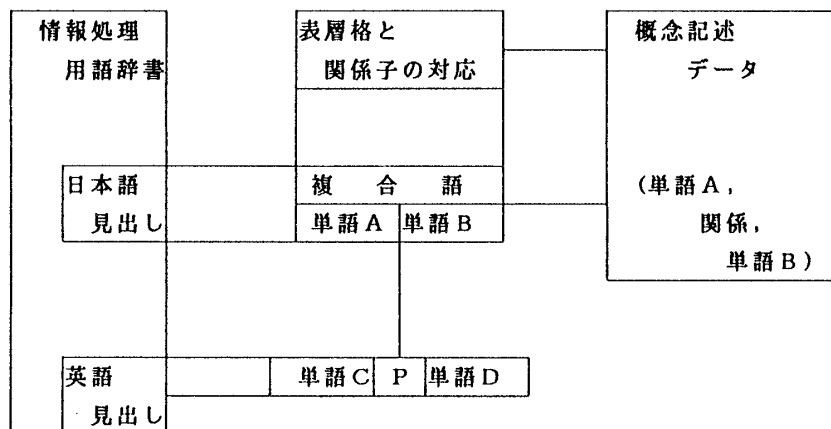
表2. 各ステップの例

概念記述の例を表3に示す。

	終点の例	方法の例	限定の例
複合語	テキスト終結	位置参照	内容検索
関係記述	(テキスト, 終点, 終結)	(位置, 方法, 参照)	(内容, 限定, 検索)

表3. 関係記述の例

実験結果のそれぞれのステップの全体的な関係を図1に示す。



P:preposition

図1. 構成語と概念記述の関係

まとめ

本発表では、情報処理用語の複合語における構成パターンの抽出、及びその構成パターンを利用した情報処理用語における複合語の概念記述の一試案について述べた。

本試案の手法では、英語で「名詞」+「前置詞」+「名詞」のパターンを取り、日本語で前置詞が訳されない複合語が少ないので、情報処理用語における適用語の範囲が狭い。しかし、複合語では以下のモデルが考えられる。

(1)「名詞句」=「名詞」+「前置詞」

(2)「Compound Word」= M3/M2/M1のとき  
M3 <r> ( M2 <r> M1 )

M:構成語

r:関係子

(1)と(2)を組み合わせると、情報処理用語における概念記述の適用語数の範囲が拡大される。

参考文献

- (1)熊野他:「専門用語辞書の試作概要」,情報処理学会第37回全国大会,(1988)
- (2)岸本他:「概念モデルに基づく概念記述の部分実験と評価」,情報処理学会第37回全国大会,(1988)
- (3)Fillmore C., The Case for Case, Reinhart and Winston(田中春美,船越道雄,『格文法の原理』,三省堂,1975)