

# 係り受け結合頻度を用いた複合名詞解析の一方法

1C-5

福田 康男 板橋 秀一

筑波大学

まえがき

複合名詞は日常よく使われる表現である。しかしこの複合名詞を、ワードプロセッサで変換すると、思いもかけない変換をしてしまうことがある。これは、複合名詞を構成する語の間の関係についてまったく考慮していないからである。

日本語解析システムにおいても、複合名詞をあまり考慮していないものが多く見られる。複合名詞を処理する場合、例えば、名詞の1回以上の繰り返しは名詞であるという規則を用いた場合、複合名詞の部分で正しくない候補が増え、その結果として、解析の精度が低下する可能性がある。また、複合名詞を全て辞書に登録するのは現実的に不可能であろう。

本報告は係り受け結合頻度を用いて複合名詞を解析することを試みたものである。

## 1. 用語

まず、本報告で使用する用語の定義を述べる。

- **名詞素**： 複合名詞を構成する短単位の名詞。名詞素には大きく分けて次の3つがある。

- a. **名詞性名詞素**： 何か1つの物や概念を表す名詞素。

例) 「名詞」 「方針」

- b. **動詞性名詞素**： 文における動詞の働きをする名詞素。

例) 「複合」 「解析」 など

- c. **形容動詞性名詞素**： 文における形容詞や形容動詞の働きをする名詞素。

例) 「高級」 「新」 など

## • 共起頻度

語どうしの係り受け関係の回数を、学習を通してカウントしたもの。この数値が高いほどその語同士の意味的な関係が強く、したがって係り受け関係が成立し易いと考えられる。

## • 係り受け結合頻度

共起頻度情報に係り受け関係の情報を加えたもの。以下、簡単化のために、単に結合頻度とも言う。

## 2. 複合名詞解析における問題

### 1. 名詞素の多品詞性

前節で、名詞素を3種に分類したが、動詞性名詞素、形容動詞性名詞素は、場合によっては、「概念を表す名詞性名詞素」として働くときもある(多品詞性)。この問題を解決するために、名詞素間の係り受け関係も共起頻度とともに辞書に登録する。したがって、辞書は図1のようになっている。

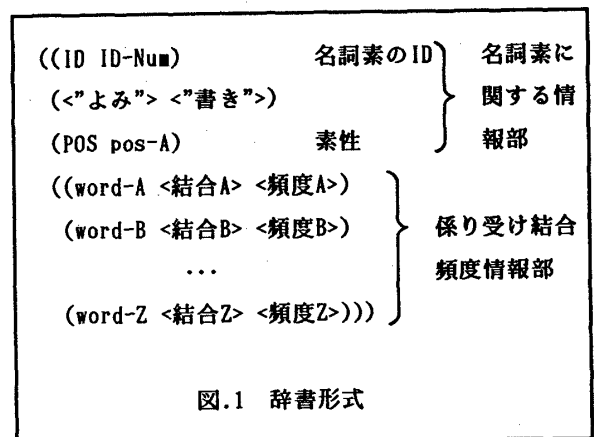


図.1 辞書形式

## 2. 未知語の問題

一般には未知語とは、辞書に登録されていない語を指す。本システムでは、このような名詞素は解析に失敗した後、ユーザーが入力して辞書に登録する。

また、この他にも、本システムでは、係り受けの関係の決定を結合頻度情報から決定するので、辞書に名詞素自体が登録されていても、係り先の名詞素との係り受け関係が辞書に登録されていない場合、その名詞素は係り受け関係が未知の語となってしまう。以下、未知語とは後者の語を指すことにする。

## 3. 解析手順

以下に解析の大まかな手順を示す。

1. 入力文字列を名詞素に切り分ける。
2. 結合頻度情報をもとに係り受け関係を決定する。  
結合頻度情報によって係り先が決定しない場合は、その係り先の決定をを保留して次の名詞素の係り受け関係を調べる。
3. 結合頻度情報によって係り先の決まらなかった名詞素の係り先を考える。係り受けに関する基本規則を満たすような係り先を仮定し、それに未知語に関する規則を適用する。
4. 全ての名詞素の係り先が決まったならば、解析結果を表示し、ユーザーに確認を求める。

以下、解析手順の各々について説明する。

### 1. 名詞素の切り分け

本システムに入力される複合名詞は分かち書きされていないので、まず最初に名詞素に切り分ける。

- 1) 共起情報を持つ名詞素とのほうが係り受け関係が成立し易いと考えられるので、まず、すでに切り分けられた名詞素の共起情報をもとに辞書を検索する。
- 2) 共起情報による切り分けに失敗した場合、最長一致法によって辞書を検索する。
- 3) 切り分けに失敗した場合、バックトラックして切り分けをやり直す。

### 2. 係り受け解析

複合名詞の係り受け解析は次の3つの規則をもとに行われる。

#### 1) 基本規則

全ての係り受けに適用される規則である。

- a) 最後の名詞素は便宜上、ダミーの"End"という名詞素に係る。
- b) 名詞素は、自分より後ろにある名詞素の1つに必ず係る。
- c) 係り受け関係は交差してはならない。

#### 2) 係り受け結合頻度に関する規則

結合頻度を用いて係り受け先を決める場合に用いる規則である。

- a) 共起頻度が高いほど係り受け関係が成立しやすい。
- b) 共起頻度が同じ名詞素が複数あった場合、より近い方の名詞素に係りやすい。
- c) 係り受けのパターンは結合頻度辞書に書かれている係り受けパターンである。

#### 3) 未知語に関する仮定に関する規則

結合頻度情報のない係り受け関係に対しては次のような規則に基づき係り受け関係を仮定して解析を行う。

- a) nnはnnに格助詞『の』の係り受け関係に係る。
- b) nnはvnに主格または目的格に係る。nnは、vnが自動詞的である時は主格、他動詞的である時、目的格であるとする。
- c) vnはnnを連体修飾する。
- d) anはnnを連体修飾する。
- e) anはvnを連用修飾する。

}	nn…名詞性名詞素
	vn…動詞性名詞素
	an…形容動詞性名詞素

#### むすび

係り受け結合頻度を用いて複合名詞を解析する方法を考えた。現在、この方法による複合名詞解析システムを構成し実験を進めている。

#### 参考文献

石崎雅人：2名詞漢字複合名詞内の名詞の意味の多義解消アルゴリズム，情報処理学会論文誌，31, 11, pp. 1696-1699(1990)