

# 日本語文書校正支援ツールの開発

## —複合名詞の統語的検定について—

3S-7

納富 一宏† 中條 和光‡ 石井 博章† 内山 明彦\*  
 †神奈川工科大学工学部 ‡福井大学教育学部 \*早稲田大学理工学部

### 1. はじめに

パソコンなど小規模システムを利用した日本語文書校正・推敲支援システムの開発を目的に研究をすすめている[1],[2].

従来、形態素解析を行なう上で問題となるのが未知語の取り扱いである。一般に、被解析テキスト中の未知語（未登録語）は、無条件でエラーとなる場合が多い。従って、多数の専門用語が出現する技術文書等の解析では、標準的な辞書では対応できず、結果として多くのエラーが発生する。これが解析精度を低下させる主要因の一つとなる。

ここで、①専門用語（未知語）の多くは複合名詞である、②複合名詞は構造を持つ、という2点に注目し、複合名詞の解析を構造と特性値というレベルで考察した。

本稿では、複合名詞の定量的評価を目的としたモデル化と解析手法について述べる。

### 2. 複合名詞モデル

#### 2.1 複合名詞の例

一般に、複合名詞は、形容詞一名詞、副詞一動詞などの係り受け関係（修飾一被修飾）を表現したものが多く、通常、転成名詞が構成要素の一部となっている。表1に例を示す。

表1. 複合名詞の例

AをBする	写真撮影, 音楽鑑賞, 現状維持, 特許申請, 海外旅行, 宇宙開発, 問題提起
AをBする人・物	新聞配達, 嘘つき, 各駅停車
AがBする	動脈硬化, 学生運動
A的なB	現代用語, 古代文明, 基本方針, 誇大妄想
AとなるようなB	問題発言
A的にBする	自動生成, 高速変換
A的にXするB	高速道路
Aを目的とするB	捜査令状, 実現方法, 読書週間
Aすることが可能なB	携帯電話, 留守番電話
AしてからBする	整列乗車
Aで(場所) Bする	現地解散, 現地集合, 海外公演, 大学教育
Aに属するB	大学教授

これらの語は一意に解釈可能な語（熟語）であるため、本来、1語として辞書に記載されるべき

だが、熟していない語でも複合化される場合があり、すべてを辞書登録で対応することはできない。

#### 2.2 名詞複合化過程—省略と転成—

複合名詞は名詞や形容詞、動詞などの自立語が複合化され、全体として1単語となった語系列として定義することができる。この複合化過程は、統語論的には文表現の簡略化を目的とした省略形式の一つである。一般に省略の対象となるのは付属語要素であり、助詞や用言の活用語尾などをあげることができる。省略に伴い品詞転成が生じる。

名詞は、意味論における概念表象であることをふまえた上で付属語の働きに注目すると、付属語は概念間の関係を示していることになる。一般的には以下のように有向グラフで表現することができる。

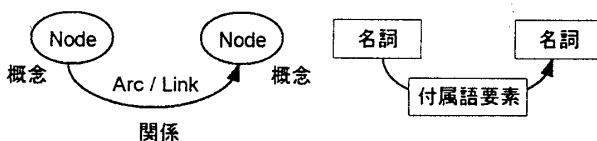


図1. 有向グラフによる表現

基本形は上図のように、2つのノードと1つのリンクで構成される2項グラフで示される。これら3つの構成要素のうち、リンク部分が省略されるためには、これを暗示する以下のような構造変形を仮定する必要がある。すなわち、「付属語要素が従属する名詞構造」である。

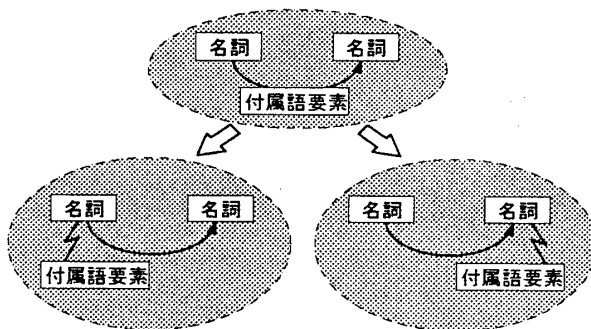


図2. 省略に伴う変形

特定の付属語要素と強い結合をもつ名詞が存在すれば、付属語要素をキーとしたクラス分類が可能となる。

構成要素の特性が大いに関係するものと予想される。

2.3 付属語従属名詞一構成要素の倒置一

複合名詞の各構成要素の倒置可否を決定するのは、名詞転成される前の品詞種別であり、それぞれの品詞連結は文法で定義されるものと考えられる。図3に構成要素の倒置と連結の例を示す。

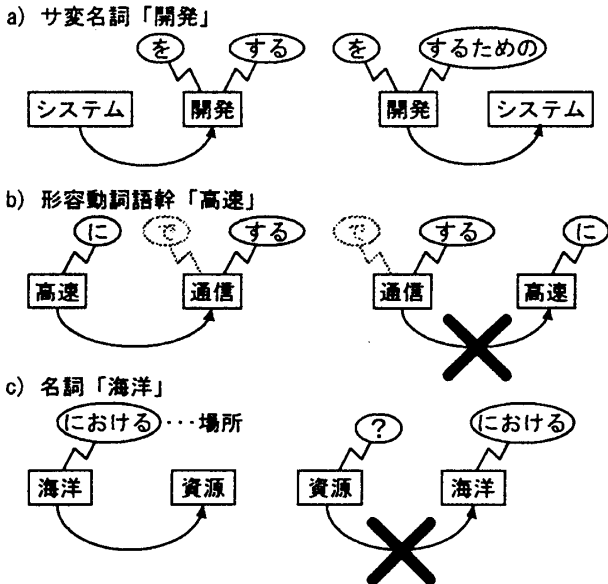


図3. 構成要素の倒置と連結

2.4 転成品詞と基本構造

構成要素の基本構造は、転成品詞別に分けると、図4の5種類となる（接頭語、接尾語を除く）。

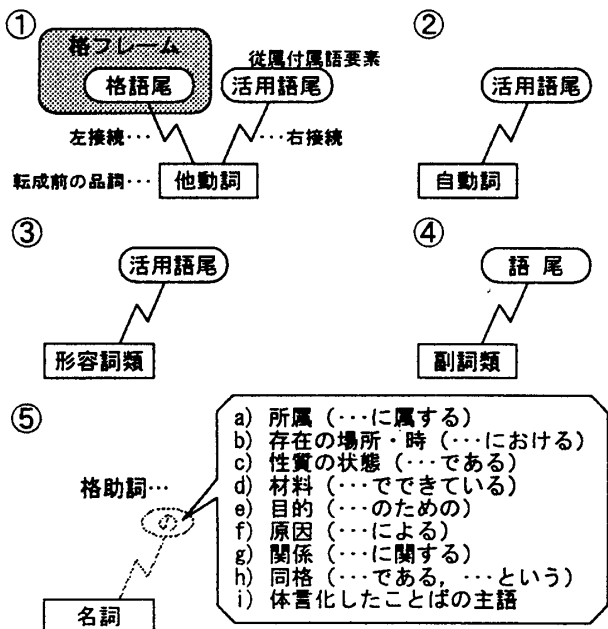


図4. 品詞による構造の違い

連結の基本形は、①名詞→他動詞、②形容詞類→名詞、③動詞類→名詞、④副詞類→動詞類、となる。3要素以上の場合も同様であると思われる。

3. 複合名詞の特性値

3.1 複合要素のマッピング

『現代用語の基礎知識'93』約7万文を対象に、非ひらがな列4文字からなる複合名詞候補57592語のうち、2文字名詞2個からなる複合名詞8432語を抽出し、前置頻度および後置頻度を求めた。

極座標系への変換を施すことで、座標値を $(r, \theta)$ にとる。これにより、それぞれ観測強度と品詞特性を表現するパラメータを取得することができる。

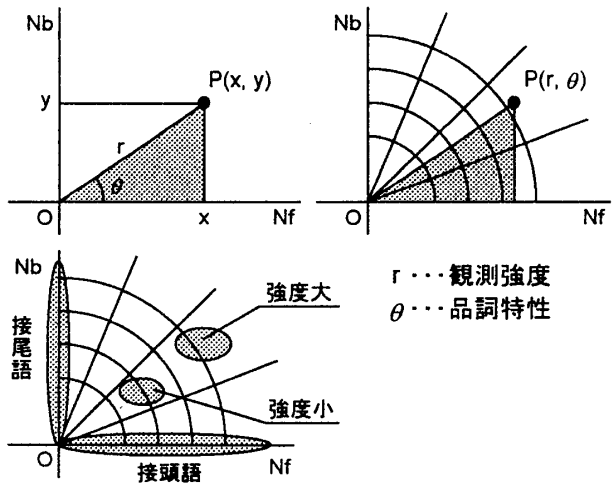


図5. 名詞のマッピング

3.2 評価実験

複合要素の前置頻度 $N_f$ および後置頻度 $N_b$ を用いて、2要素 $w_1, w_2$ の連結評価関数 $f(s)$ を次式で定義する。

$$s \Rightarrow w_1 + w_2$$

$$P_f(s) = \frac{N_{f1}}{N_{f1} + N_{f2}}, \quad P_b(s) = \frac{N_{b2}}{N_{b1} + N_{b2}}$$

$$f(s) = \frac{1}{2} \{ P_f(s) + P_b(s) \}$$

大学生15名による評価実験を行なった。4字複合語8432語からランダムに200語を選び、うち100語について前後を入れ換えた。これら200語を提示し、複合名詞としてあり得るか否かを7段階で評価してもらった。結果として、評価関数と人間による判定との相関計数は、0.78であった。

4. まとめ

複合名詞のモデル化と解析手法について述べた。現在、ツールとしての有効性評価を行なっている。今後、未知語学習について検討する予定である。

文献

[1] 納富, 内山他: 日本語文書校正支援ツールの開発一解析手法の検討と評価一, 第46情処全大, (1993.03).  
 [2] 納富, 内山他: 日本語文書校正支援ツールの開発一動詞格フレームと名詞ソーラスの利用一, 第47情処全大, (1994.10).