

語彙概念構造を利用した複合名詞内の係り関係の解析

竹内 孔[†] 内山 清子^{††} 吉岡 真治^{†††}
影浦 峽[†] 小山 照夫[†]

本研究では、主辞がサ変名詞である複合名詞の語構成において、構成する単語間の係り関係を支配する語彙的性質に着目し、それに基づく複合名詞解析モデルの作成を試みる。主辞がサ変名詞の複合名詞内の係り関係の解析は、並列関係の場合を除くと、主辞であるサ変名詞の項関係なのか修飾なのかを同定することが解析の第1歩である。項関係とは名詞が動詞の目的語や主語といった関係であることを意味している。本論文では、この関係をとらえる方法として、語彙概念構造を利用した動詞の分類と、その構造を利用した名詞の分類に基づく複合名詞解析手法を提案する。情報処理関連の専門用語と新聞記事中の一般的な複合名詞に対してテスト的な実験を行った。その結果、平均で1231語の複合名詞に対して約99.4%の複合名詞を正しく解析する結果を得た。

Analysis of Japanese Deverbal Compounds Based on Lexical Conceptual Structure

KOICHI TAKEUCHI,[†] KIYOKO UCHIYAMA,^{††} MASAHARU YOSHIOKA,^{†††}
KYO KAGEURA[†] and TERUO KOYAMA[†]

In this paper, we describe a principled approach for analyzing relations between constituent words of compound nouns, specifically those whose heads are deverbal nouns, based on the classification of deverbal nouns by their lexical conceptual structure (LCS) and the classification of nouns in modifier position vis-a-vis LCS of head deverbal nouns. There are two kinds of relations of compounds with head deverbal nouns. The one is that a modifier noun becomes an argument of deverbal head and the other is that a modifier becomes an adjunct. It is an important starting point for analyzing relations to disambiguate the two kinds of relations. Through the qualitative analysis of the data and the experimental evaluation of 1231 compound nouns, we show that the use of LCS as the theoretical basis is very promising for constructing compound analyzer.

1. はじめに

複合名詞は様々な分野において、中心的な概念を表記するために用いられており、文書の機械処理を展望する場合、その解析手法の確立が重要な課題となる。

コンピュータで言語処理を行うという観点から複合名詞解析は大きく2段階に分けることができる。すなわち、

- 1) 語構成単位に分割し、
 - 2) その係り関係を明らかにする。
- 1)の語構成単位の分割には、1.1) 最小単位の決定の問

題と1.2) 複合名詞内の区切りの曖昧性解消の問題がある。1.1の問題は言語学の視点からは奥深い問題²²⁾であるが、当面は最小単位となる単語集合が辞書として与えられていると仮定することにより回避することができる。1.2の問題は区切り方の異なる複合名詞どうしを評価する手法が確立してないことから、今回は検討の範囲外とし、本研究では、これらの処理は完了していると仮定する。

本論文では複合名詞解析の第1段階として、2)の係り関係の解析を目標とする。従来、工学的な複合名詞解析のアプローチとして、主に確率的な手法による^{3),6),7),18)}もの、独自に設定した規則ベースによる^{9),12)}ものがある。しかし、確率的な手法は解析対象に応じた分野に対して、高い質の学習データ²⁾が必

たとえば「購入-金」が「購-入金」なのかどちらの分割がよりその文脈で適しているかを比較すること。ここで「-」記号は単語の切れ目を表す。文献4)参照。

[†] 国立情報学研究所

National Institute of Informatics

^{††} 慶応義塾大学大学院政策・メディア研究科

Graduate School of Media and Governance, Keio University

^{†††} 北海道大学大学院工学研究科

Graduate School of Engineering, Hokkaido University

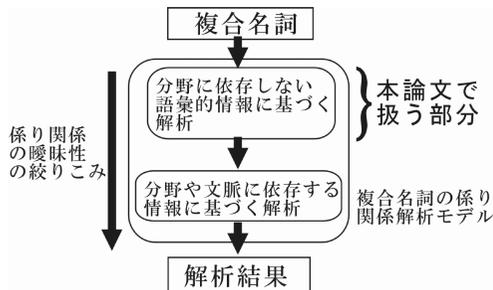


図1 複合名詞内の係り関係解析モデルの全体図

Fig.1 Overview of our model for analyzing relations between constituent words of compound nouns.

要となるため、その作成に多大な時間と労力を必要とする。また、これまで行われてきた規則ベースの適用を試みるアプローチの多くでは適用した問題に特化していることが多く、他の問題に応用するのが容易ではない。さらに、意味付与^{5),10),11),20)}による手法もあるが、意味記述が係り関係の解析にどのように結び付くのが明らかではないため、係り関係の解析モデルに利用することが難しい。

これに対して、近年、文法理論に基づく複合名詞の記述分析^{14),15)}が進んでおり、各語の分野に依存しない性質に基づく語構成の仕組みが明らかになりつつある。本論文ではこの語彙的性質に注目した複合名詞内の係り関係を解析するモデルの作成を試みる。

解析モデル全体として、分野に依存しない語彙的情報に基づく解析の後に、分野や文脈に依存する情報に基づく解析部分があると仮定(図1参照)する。分野に依存しない解析で曖昧性を絞り込み、次に分野に依存する解析に委ねることで拡張性、応用性のあるモデルの構築につながると考える。

対象とする複合名詞は言語学的検討が多くなされている動詞由来複合語とする。動詞由来複合語とは主辞がサ変名詞の複合名詞のことをいう。語彙的性質として、語彙概念構造(Lexical Conceptual Structure: 以下、LCSと表記する)をもとにしたカテゴリを提案する。この語彙概念構造をTLCSと表記する。TLCSを単語に付与することで、2単語からなる複合名詞の係り関係を解析するモデルが作成できることを明らかにする。

提案する規則に基づく複合名詞解析モデルを評価するために、解析実験を行った。対象は、情報処理用語

辞典に記載されている2単語からなる動詞由来複合語で、基本単語はすべて辞書に登録した。項関係が修飾かの係り関係の解析により816語中99.4%の語を正確に解析した。さらに、日経新聞についても同じ手法を適用し、同様に高い精度を得た。これらの結果から、解析手法に対する考察を行い、本手法の発展方向について考察する。

以下、2章では、語彙概念構造を用いた基本的な解析モデルの方針、3章ではTLCSの説明と付与方針、4章ではTLCSをもとにした名詞のカテゴリ分け、5章では専門用語に対して行った解析実験の結果を各々提示し、6章では新聞記事中の専門用語に対する提案手法の見通しを提示し、7章で総括を述べる。

2. 語彙概念構造を用いた複合名詞解析モデルの枠組み

本論文が用いる語彙概念構造は文献1)が提案した項構造という動詞の意味表記をもとにしている。これは概念構造において、動詞が名詞に与える意味的な役割を表しており、外項を動作主(Agent)、内項を動作記述の対象(Theme)または目的(Goal)とした。たとえば「操作」というサ変名詞の項構造は操作する主体をAgent、操作される対象をThemeとすると

- 操作 (Agent (Theme))

と表現する。⟨ ⟩ は内項を表す。内項と外項を導入できると、文献14)から、外項となる名詞は複合名詞とはならないという制約を利用できる。

作業員が機械を操作する → 作業員の機械操作

*機械の作業員操作

下段の文では作業員が機械を操作するという意味にはならない。この内項と外項は項構造上の表現であるが、これらは、文中の統語構造において、主格、対格、与格として出現する。内項、外項とこれらの格との対応関係は動詞の語彙概念構造のタイプにより異なる。3章で説明する。

「機械操作」のような内項と結び付いた複合名詞の関係を項関係(または内項関係)と呼び、一方、項関係以外の関係は修飾関係として扱う。内項を中心に語彙概念構造に拡張したモデル化を行う。

2.1 係り関係の整理

形態論における文法的知見^{14),15)}から、動詞由来複合語では、前項の名詞と主辞のサ変名詞との関係は大きく項関係か修飾かの2つに、まず、分けること

LCSは、基本的に「枠組み」および「考え方」を指すが、LCSに基づく具体的な記述体系を指す場合もある。混乱を避けるために我々のLCSに基づく分類体系を、第1著者の頭文字をとってTLCSと呼ぶことにする。

詳細な説明は文献17)を参照。*印は目的とする意味をなさないという意味を表す。

ただし「振幅 位相 変調」のような並列関係を含む語は除く。

ができる。たとえば、

- 機械 翻訳
- 機械 操作

の場合、上段は「機械」が主辞の「翻訳」を修飾する解釈であり、下段の例では通常、「機械を操作」という内項の解釈となる。修飾関係の内訳には道具、様体、など様々な要素があると思われるが、係り関係を解析するはじめの段階として、項関係か修飾関係かをまず分類することを本論文では目的とする。

さて、この係り関係の違いは主辞の動詞の違いから生じていると考えられる。そこで、動詞の意味を詳細に分類することができる LCS をもとにした動詞の TLCS (3 章を参照) を利用して、これら複合名詞をとらえ直すことを試みる。

2.2 TLCS による複合名詞解析モデル

TLCS 表記を用いると、上例の主辞はそれぞれ、

翻訳 [x CONTROL [BECOME [y BE AT z]]] (1)

操作 [x ACT ON y] (2)

となる。‘x’ は上述の動作主 (Agent), ‘y’ は対象 (Theme), ‘z’ は目的 (Goal) を表し, ‘ACT’, ‘CONTROL’, ‘ON’, ‘BECOME’, ‘BE’, ‘AT’ は基本的な意味述語を表している (詳細は 3 章で説明する)。「機械翻訳」の場合、修飾関係なので「機械」は目的語 ‘y’ ではなく、「翻訳」の意味述語を修飾するととらえる。一方、内項関係の「機械操作」の場合は「機械」は式 (2) の ‘ON’ の項の ‘y’ に代入されるととらえる (図 2 を参照)。

本研究では、この内項か修飾関係かが TLCS の要素に依存して決まると仮定して、TLCS 内の ‘y’ に対する意味的な操作に着目した解析モデルを構築する。

たとえば「翻訳」の TLCS では目的語 ‘y’ は ‘BECOME’ に支配される項となっているが「機械」はこの型の内項 ‘y’ に入らないと分類しておく。すると、「翻訳」と同じ TLCS を持つ動詞「処理」「学習」に対して「機械」は同様に修飾関係であることがうまくとらえられる。一方「操作」の TLCS の ‘ON’ に対して「機械」は内項になると分類すると、同じ TLCS を持つ「制御」「運転」に対して「機械」は内項関係となることが表現できる。このように TLCS を動詞に付与しておき、意味述語に対する関係を体系的に整備し、一貫して付与しておくことで、内項関係か修飾関係かを解析するモデルが構築できることを本論文では明らかにする。

ただし、この解析では、TLCS の項の構造に対して

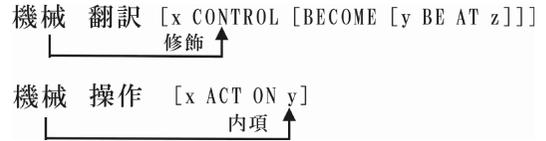


図 2 「機械」の係り関係の違いをサ変名詞の TLCS でとらえる
Fig. 2 Disambiguating relations using TLCS types of deverbal heads.

非内項か否かを扱っているだけで、修飾関係に対して何の制約を課すこともできない。つまり、我々の解析モデルでは非内項関係と判定すれば、修飾関係を意味するが、内項関係と判定しても、修飾関係の可能性も残る。例をあげて説明すると、

- 異常発生

は、文脈によって内項関係である「異常が発生」という解釈と修飾関係の「異常に発生」という解釈がある。提案するモデルでは、この場合、内項関係になりうるかと解析するにとどめ、文脈に依存するモデルにさらなる解析を委ねることとする。つまり、本提案手法では、非内項となる場合の規則を重視した複合名詞解析モデルの構築を行う。

3. 複合名詞解析に用いる語彙概念構造

3.1 基本構造

LCS は動詞の意味を項構造と意味述語に分解して表現する方法で形態論で研究されている。すべての動詞を表すことのできる LCS の組が確定しているわけではないが、文献 15), 16) を参考に複合名詞解析のためのサ変名詞に付与する TLCS の基本構造を作成した。

- 1 [x ACT ON y] 操作, 測定
- 2 [x CONTROL[BECOME [y BE AT z]]] 処理, 翻訳
- 3 [x CONTROL[BECOME [y NOT BE AT z]]] 遮へい, 抑止
- 4 [x CONTROL [y MOVE TO z]] 伝送, 伝搬
- 5 [x=y CONTROL[BECOME [y BE AT z]]] 回復, 終了
- 6 [BECOME[y BE AT z]] 飽和, 分布
- 7 [y MOVE TO z] 遷移, 移動
- 8 [x CONTROL[y BE AT z]] 継続, 維持
- 9 [x CONTROL[BECOME[x BE WITH y]]] 認識, 予測

文献 21) では修飾関係にいて拡張した LCS を仮定してモデルを提案している。

文脈はなく、独立して取り出した場合の解釈である。

- 10 [y BE AT z] 関係, 位置
 11 [x ACT] 会議, 行列
 12 [x CONTROL[BECOME [[FILLED]y BE AT z]]] 署名

本報告で使用する TLCS は上記の 12 種とする。5 章の解析実験で出現するすべてのサ変名詞はこのどれかの TLCS に分類される。‘CONTROL’, ‘BECOME’, ‘MOVE’, ‘TO’, ‘ACT’, ‘ON’, ‘BE’, ‘AT’, ‘WITH’, ‘FILLED’, ‘NOT’ は意味述語を表す。また, ‘x’ は外項を表し, ‘y’ と ‘z’ は内項である。これらは, 名詞が入る変項(variable)を示す。項関係と格との間には, 動詞の TLCS タイプと密着して, 明確な対応関係がある。自動詞の中で TLCS 6, 7, 10 タイプの動詞では ‘y’, ‘z’ が各々, 主格, 与格と対応するが, 他の TLCS タイプに属する動詞では, ‘x’, ‘y’, ‘z’ が各々, 主格, 対格, 与格と対応する。この 3 つの項のうち, 複合名詞において項を取り込む ‘y’ に注目して TLCS に対する分類を行う。

TLCS の意味述語には制約関係が存在する。BNF を利用して記述すると以下ようになる。

```

<VERB> ::= <BE> | <MOVE> | <BECOME>
          | <CONTROL> | <ACT>
<BE> ::= BE AT | NOT BE AT | BE WITH
<MOVE> ::= MOVE TO
<BECOME> ::= BECOME <BE>
<CONTROL> ::= CONTROL <BE>
              | CONTROL <MOVE>
              | CONTROL <BECOME>
<ACT> ::= ACT | ACT ON
  
```

<VERB> が生成するすべての動詞を表している。この制約では ‘CONTROL [BE WITH]’ のような 12 種の TLCS に存在しないパターンも生成するが, 意味的な制約から TLCS に含まれていない。また, 動詞の種類によっては項の在り方に制約が生じる場合があり, このような制約を ‘=y’ や ‘FILLED’ で表している。これら制約をもとにサ変名詞に対して調べた結果が上記の 12 種類 TLCS である。

提案する TLCS はいくつかの意味述語が組み合わさってできているので, 以下では, まず, 構成要素について簡易な説明を与え, 全 TLCS タイプに対する横断的な性質を説明する。次に, 動詞に対する概略的な TLCS の付与法を説明する。各意味述語に対する

詳細は文献 14), 15) を参照されたい。

3.2 TLCS 内の意味構成要素の説明

- ‘y BE AT z’ は ‘y’ がある状態 ‘z’ に存在することを表す。
- ‘BECOME’ はこの述語の右側に記述される述語の状態に変化することを表す。
- ‘y MOVE TO z’ は ‘y’ が ‘z’ という場所に移動することを表す。これは状態変化を表す構成要素 ‘BECOME[y BE AT z]’ をさらに移動という観点から, 細分類したものである。移動動詞では, 「道のり」の名詞をヲ格で取り込むことができる特徴がある。たとえば「状態間遷移」は「状態間を遷移する」のように, ヲ格の関係となる。この特徴から, ‘MOVE TO’ を細分類した。
- ‘x CONTROL’ は ‘x’ がこの右側に記述される述語の変化をコントロールすることを表す。
- ‘x=y’ は内項 ‘y’ 自身が主体的な意味を持っていることを表しており, 自他交替現象が見られる。これは, たとえば「負荷が分散する」「負荷を分散する」のように「れる」などの助動詞を必要とせず主語と目的語が入れ替わる現象である。これは能格動詞と呼ばれ, 漢語動詞に観測される。この ‘x=y’ の性質を持つのは TLCS 番号 5 のみである。
- ‘x ACT ON y’ は継続的な動作で, ‘x’ が ‘y’ に対して直接的な動作を与えることを表す。
- ‘x ACT’ は ‘x’ の継続的な動作を表す。
- ‘x BE WITH y’ は ‘x’ が ‘y’ と一体化して存在することを表す。
- ‘NOT’ はこの右側に記述される述語の意味を否定することを表す。
- ‘[FILLED]y’ は ‘y’ の項が埋まっていて, 複合名詞において, 項をとらないことを表す。

3.3 TLCS に対する横断的な説明

‘BECOME’ と ‘MOVE TO’ は状態変化を意味する述語であるため, これを含む TLCS 番号の 2~7, 9 は完了性を持つ。たとえば, TLCS 番号 6 の動詞「飽和」に対して「している」を付与すると「飽和している」のように動作完了の意味が現れる。一方, ‘ACT’ は継続動作を表すので, TLCS 番号 1 と 11 の動詞には完了性はない。たとえば, 1 の動詞では「操作している」のように今行っている動作の意味となる。

‘ACT ON’ が ‘CONTROL’ が TLCS にある場合, 外項の ‘x’ と内項の ‘y’ があるため, 他動詞となる。こ

外項 x は複合の際に目的語を取り込むことはない(外項排他的制約¹⁴⁾)。また, 内項 ‘z’ は複合の際, 補語を取り込むことがあるが, 動詞が限られているので別途扱う。例)「応用依存」は「応用に依存する」。

英語の他動詞にも観測される。「break」など。

表1 TLCSのカテゴリ分け
Table 1 Typology of TLCS.

動詞の種類	完了性	
	あり	なし
他動詞	2,3,4 9	1 8
非対格自動詞	6,7	10
非能格自動詞	12	11
能格動詞	5	

の場合の内項 'y' は文中ではヲ格をともなって出現する (TLCS の 1 ~ 5 , 8 , 9 が該当する) . 逆に , 'x' と 'y' のどちらかが TLCS 内にはない動詞は自動詞となる . ガ格の名詞が内項の場合は非対格自動詞 (TLCS の 6 , 7 , 10) , ガ格の名詞が外項の場合は非能格自動詞 (TLCS の 11 , 12) , 自動詞と他動詞の両方の性質を持つのが能格動詞 (TLCS の 5) である .

3.4 動詞に対する TLCS の付与

動詞に対して用意した 12 種の TLCS を付与する手順について説明する . 手がかりは , 1) 動詞の格のとり方による他動詞 , 自動詞の分類と 2) 動詞の完了性である . これらを軸とした TLCS 分類を表 1 に示す .

ここで , 表中の番号は 3.1 節で示した TLCS 番号である .

1. 動詞の格のとり方を調べる .

対象とする動詞のガ格 , ヲ格 , 二格のとり方を調べる . ヲ格をとる場合 , 他動詞か能格動詞である . 自他交替を起こすものは , 能格動詞とし , 残りが他動詞である .

ヲ格がない場合は自動詞となる . 前節で記述したように , ガ格の名詞が内項か外項かを意味的に考えることで , 非対格自動詞か非能格自動詞に分類する .

2. 動作の完了性を調べる .

統語的テストとして , 対象とする動詞に「している」を付加してその解釈から以下のように分類する .

- 完了状態の解釈ができる場合 , その動詞の TLCS は 2 ~ 4 , 6 ~ 7 , 9 , 12 となる .
- 完了の解釈がない場合 , その動詞の TLCS は 1 , 8 , 10 , 11 となる .

3. 各個別の分類を意味的な判断で行う .

3-1 他動詞で完了性のある動詞 (表 1 中の左上)

- 移動の意味の有無を調べる . テストとして , 二格の名詞を動詞に付与して移動の意味が認識される場合 ¹ TLCS の 4 に分

類する .

- 変化する対象を調べる . ヲ格をとる名詞が変化するのではなく , 主体の知覚や考えなどが変化する場合 TLCS の 9 に分類する .
- TLCS2 と 3 の違いは述語 'NOT' の有無である . 否定的な動作がある動詞は TLCS の 3 に分類し , そうでない動詞は 2 に分類する .

3-2 他動詞で完了性がない動詞 (表 1 中の右上)

主体の動作性の意味の強いものが TLCS の 1 である . テストとして動詞に「たくさん ~ している」を付与した場合 , 主体の動作量が多い場合 ² は TLCS の 1 に , 結果できたものの量を指す場合 ³ は TLCS の 8 に分類する .

3-3 非対格自動詞で完了性がある動詞 (表 1 中の左 , 上から 2 行目)

上記でも記述したように , 移動の意味の有無を調べる . 二格の名詞を付与して移動の意味がある場合 ⁴ は TLCS の 7 に分類し , ない場合は 6 に分類する .

4. 名詞の分類

2 章で述べたように , 動詞由来複合語の係り関係は主辞の TLCS の内項 'y' に前項の名詞が入れば , 内項関係で , そうでなければ , 修飾関係である . 本研究では , 名詞が内項に入るか否かは TLCS 中の 'y' の意味述語に対する名詞の性質によると仮定する . すると , 名詞の分類は TLCS 意味述語の操作を受ける (内項になる) か受けないか (修飾関係になる) の 2 値の属性値で表すことができる .

図 2 では「機械」が TLCS 番号 2 の内項には入らない例をあげた . 3.1 節に示したように動詞の TLCS は 12 種類あるため , 単純に考えると 1 つの名詞に対して , 12 種類の TLCS について内項か非内項の性質かを記述する必要がある . しかし , TLCS の内項 y の意味述語のうち , 似ている構造をまとめることで , 少数の属性で特徴をとらえることができる . こうした , 操作の明確性が TLCS を用いる本手法の利点である . 本論文では , 5 つの属性に絞ら込んだ . 次節では , それぞれの属性と現象について説明し , その次の節で動詞の TLCS と名詞の分類から係り関係に対する結び付きについて示す .

² 「たくさん操作している」は操作量が多い .

³ 「たくさん維持している」は維持しているものが多い .

⁴ 「公園に移動する」

¹ 「データを相手に伝送」

名詞の分類には、基準となる TLCS が付与された動詞が必要となるがそのような辞書はまだ存在しない。そこで、情報処理用語辞典⁸⁾と新聞記事から取り出した 247 語のサ変名詞に TLCS を付与して参考にした。この分類付与の工程において、事例の確認として日経新聞の産業・金融・流通新聞 92～98 年の全文記事、角川書店の国語辞典を参考にした。

4.1 属性による名詞の分類

以下の 5 つの属性について名詞の性質を調べ、内項になる可能性がある(以下, (+) で表示)かない(以下, (-) で表示)かを人手で分類する。注目した属性について説明し、非内項となる例をあげる。

属性 1: 格のとり方に関する分類

文中である単語がガ格、ヲ格をとって、主語、もしくは目的語となることがない場合、‘-GAO’ に分類する。

文献 17), 19) の考察から文中でガ格やヲ格をとらない単語が存在する。形態素解析システム¹³⁾などでは、名詞に分類されている場合もあるが、これらの単語は複合語中にしか現れず、どの TLCS を持つ動詞に対しても内項にならない。‘-GAO’ の例として「積極」「消極」「遠隔」などがあげられる。

属性 2: TLCS の ‘ON’ に対する分類

TLCS 番号 1 の ‘ON y’ の形をとる動詞の内項として文章中に現れない名詞に対して ‘-ON’ として分類する。

意味述語 ‘ACT ON y’ は ‘y’ に対する直接的な働きかけを意味している。図 2 にもあったように「機械」のような実体のあるものは内項になる。しかし、実体のない「温度」も言語上では「温度測定」「温度制御」のように内項となり、直接的な働きかけを受けることができる。よって、これらの分類を行うのに従来の概念辞書などが利用できないことが分かる。‘-ON’ とする単語は「平行」「連鎖」などである。

属性 3: 外的コントロールに対する分類

TLCS のタイプが 2, 3, 4 の ‘x CONTROL’ の形をとる動詞の内項として文章中に現れない名詞に対して ‘-EC’ として分類する。

TLCS 番号 2, 3, 4 に属する動詞は内項 ‘y’ が外項 ‘x’ の影響を受けて、状態変化を起こすことを示している。本論文ではこれを外的コントロール性と呼ぶことにする。意味述語 ‘MOVE TO’ は移動を意味しているが、状態変化という観点では ‘BECOME’ と同じである。‘-EC’ は外的コントロール性を持たないことを示しており、例として「予測」「集団」などがある。

属性 4: 内在的コントロールの TLCS に対する分類

表 2 名詞と TLCS の組合せによる係り関係

Table 2 Categorization of combination of modifier nouns and TLCS of deverbal heads.

係り関係	名詞の分類	TLCS パターン
修飾関係	-GAO	any
	-ON	1
	-EC	2,3,4
	-IC	5
	-UA	6,7
	any	10,11,12
項関係	+GAO	8,9
	+ON	1
	+EC	2,3,4
	+IC	5
	+UA	6,7

TLCS 番号 5 のタイプの動詞の内項として文章中に現れない名詞に対して ‘-IC’ として分類する。

5 の TLCS は内項が同時に外項の役割を持つことを表している。名詞が自分自身状態変化することを示している。これを内在的コントロール性と呼ぶことにする。似た性質として、6 の TLCS があるが、次の属性 5 として別に分けた。たとえば、TLCS 番号 5 に属する動詞「終了」に対して「会議終了」というが「*梅雨終了」とはいえない。しかし、TLCS 番号 6 に属する動詞「明ける」に対して「会議明け」「梅雨明け」とどちらも内項となる。このように TLCS 番号 6 は自然発生的現象を含むので、扱いを変えている。‘-IC’ はこの内在的コントロール性を持たないことを表しており、例としては「統計」「線形」などがある。

属性 5: 非対格自動詞に対する分類

TLCS が 6, 7 の内項として文章中に現れない名詞に対して ‘-UA’ として分類する。

TLCS 番号 6, 7 の動詞は非対格自動詞と呼ばれるもので、内項 ‘y’ が自ら状態変化を引き起こすことを表している。属性 4 との違いは上述のとおり。‘-UA’ として分類される例は「順序」や「並行」などである。以上 5 つの属性が名詞の分類であるが、上記で対象としていない TLCS (8~12) が残っている。これらについては、対象とする名詞カテゴリは必要ない。次の組合せの節で説明する。

4.2 名詞と TLCS の組合せによる係り関係

名詞の分類と TLCS の組合せにより決まる係り関係を表 2 に記述する。

TLCS のパターンの番号は 3.1 節の TLCS 番号で

*印は非文を表す。また、非文になるものは ‘-IC’ とするのではなく、修飾関係ではないので対象外とする。

ある．表中の any は分類に依存しないことを意味している．

係り関係を解析するモデルは，名詞と主辞の動詞のカテゴリを調べ，この表を参照する．上記の表の上段では修飾関係になる組合せを示しており，どれか1つでも組合せにあてはまれば，その関係は修飾となる．どれもあてはまらなければ，項関係として解析する．

TLCS 番号 10, 11, 12 の構造はどのような名詞がきて項関係となる複合語の構造にならない．11 は自明，10 と 12 は今回調べた範囲から判断した．また，TLCS 番号の 9 の動詞では，ガ格，ヲ格をとる名詞（属性+GAO）が相手の場合に項関係になる．これは TLCS の 9 に含まれる 'x BE WITH y' において外項 x 自身が変化し，内項 y に対して特別な状態変化，あるいは動作対象としないことと対応している．また，TLCS の 8 も同様に，項関係となるが，このタイプの TLCS も 'BECOME' 項がないことにより，作用対象に対して変化を求めないことが制約を生まないことに対応している．

5. 専門用語の解析実験と考察

5.1 解析モデルと実験結果

情報処理用語辞典⁸⁾から取り出した 2 単語からなる動詞由来複合語 816 語について係り関係の解析実験を行う．前項の名詞と主辞のサ変名詞に対して，TLCS による分類を行っておく．複合名詞解析モデルはこの分類と表 2 に基づいて解析を行う．入力された複合名詞に対して以下のような手順で解析する．

- step 1 主辞の TLCS が 3.1 節の 10, 11, 12 の場合，修飾関係と判定し終了する．そうでなければ，次に進む．
- step 2 前項の名詞の属性が '-GAO' ならば，修飾関係と判定し終了する．そうでなければ，次に進む．
- step 3 表 2 の '-ON', '-EC', '-IC', '-UA' の各規則を比較し，適用できれば修飾関係と判定し終了する．そうでなければ，次に進む．
- step 4 項関係と判断して，終了する．

評価を行うために，専門用語辞典の情報をもとに，816 語に対して，内項関係か修飾関係かを人手で解析し，解析モデルが出した結果と比較する．評価データの性質は，主辞のサ変名詞の異なり数が 208 種類，名

理由は明らかでないが，10 は漢語動詞に限る．和語の場合，TLCS 10 に対応する動詞としては「ある」が存在するが，これは，項関係となる複合名詞を構成する．たとえば「教師あり学習」という専門用語は「教師がある」学習方法という解釈ができる．つまり「教師」は「ある」の内項である．

表 3 正しく解析した中で適用されたルールの頻度
Table 3 Statistics of effective rules applied to analysis correctly.

手順	適用したルール	頻度
step 1	TLCS が 10,11,12 の場合	45
step 2	-GAO	219
step 3	step 3 の合計	270
	-ON	65
	-EC	142
	-IC	20
	-UA	43
step 4	項関係	277
	合計	811

詞の異なり数は 492 種類であった．このデータに対する解析実験で，提案する解析モデルは 99.4% の専門用語について項関係か修飾かを正しく判定した．

正しく解析した 811 語のデータ中で，上記のどの手順におけるルールがどれだけ適用されたかを表 3 に示す．単独の属性では，名詞属性 1 の格のとり方に関するルールが有効に働いており，step でみると，TLCS の意味述語に関するルール，step 3 (属性 2 から 5) が有効に働いている．

この結果は我々の方法が有効であることを示しているが，2 回以上出現する名詞が 816 語中，484 語（約 60% 程度）であったこと，また，分野が情報処理に関する専門用語に限定されていることから，参考的な値としてとらえる．

5.2 考察

上記の実験では，816 語中 5 語について解析が失敗した．失敗の理由は大きく分けて，3 つに分類できる．以下，順に説明する．まず，表 2 で設定した TLCS パターンとの組合せのルールで例外となり，正しくとけなかった複合名詞である．

磁気遮へい
磁気記憶

上段は項関係，下段は修飾関係となる．辞典に記述されている意味から「磁気」を「記憶」することはないという解釈を正解とした「遮へい」「記憶」の TLCS は 3 と 2 でともに 'CONTROL' と 'BECOME' を含む他動詞である．表 2 で示したようにこれらの TLCS に属する動詞は同じ振舞いを行うとしてまとめているため，誤りとなった．この「遮へい」と「記憶」の TLCS は意味述語 'NOT' の有無の違いがあり，この述語が係り関係に一定の影響を与えている可能性もある．今後，さらなる例を調べて，検討したい．

次に示すのは，名詞の分類と TLCS のタイプがまったく同じであるにもかかわらず，係り関係が異なる場

合である。

{	論理設計	{	入力命令
{	論理表現	{	入力導出

上段は「論理(回路)を設計する」「入力を命令する」という解釈で、どちらも項関係である。下段は「論理的な表現を行う」「入力から導出する」という意味で修飾関係である。主辞の「設計」「表現」「命令」「導出」はすべて同じ TLCS タイプの 2 に分類され「論理」と「入力」はともに、上記サ変名詞に対して、修飾関係であると分類されるので、上段の項関係を正しく解析することができなかつた。これらの複合名詞はサ変名詞の TLCS による分類がまったく同じであることから、提案する枠組みではうまくとけない。多くの事例を検証することで、TLCS のパターンに対する考察を行いたい。

最後に示すのは、係り関係の解釈が微妙で、正解判定にゆれが原因で誤りとなった可能性のある複合名詞である。

{	数値解析	{	数値計算
{	数値せん孔	{	数値制御

上段は項関係を正解とし、下段は修飾関係を正解とした「解析」と「せん孔」は TLCS の 2 「計算」と「制御」は TLCS の 1 に分類されるため、同じ TLCS の主辞に対して「数値」の係り関係が異なっている。この点では、前述の誤りパターンと同じである。しかし、上段の複合名詞において、「数値を利用した解析」のように修飾関係とする解釈も強く感じられるため、正解判定のゆれのために誤りとなった可能性もある。今後、こうした微妙な関係に対して、いい回しによるテストで判定できる方法を構築する必要がある。

また、正解と判断された係り関係の中にも、分野が異なることによって、項関係か修飾関係かの判定の部分で、ゆれが出る可能性のものもある。

● 障害 保守

は内項関係として「障害を保守する」を正解とした。「保守」は TLCS の 8 に分類されるため、表 2 からガ格やヲ格をとる名詞に対して内項関係となり、判定と一致する。しかしながら、分野が異なれば「障害から保守する」のように、修飾関係に解釈されることもあるかもしれない。その場合でも、この枠組みでは、内項関係と解釈するとどめ、分野や文脈に依存するモデルの中で扱いを検討することになる。

表 3 では、TLCS のタイプとは関係ない名詞属性 1 (表中の step 2) が係り関係の決定に有効に働いている

が、その他の修飾関係を決定する要因は TLCS と名詞の分類による。特に、名詞の分類に関係なく、TLCS のタイプだけで修飾関係と決定される step 1 で 45 語解析しており、TLCS に基づく分類だけでも効果的に効く事例が多いことを示している。

また、TLCS に対応した名詞分類である属性 2~5 に対するルール(表 3 中の step 3 の部分)は頻度が 270 語と修飾関係のほとんどを決定しており重要な属性であることが分かる。しかしながら、モデルの拡張性という観点からこれらの属性が、複合名詞解析以外でどのような意味を持つかがまだ明らかではない。これは今後の課題としたい。

本モデルで内項関係と出力するのは、非内項となるルールに抵触しなかつた場合であるが、表 3 から 277 語正しい解析であった。次ステップの解析の観点から、本モデルは内項関係に対してはかなり見通しが良い。たとえば、動詞の TLCS が明らかであることから、係り関係をガ格、ヲ格、ニ格を用いた格関係として表現することが容易である。

一方、修飾関係は同表から 534 語と、内項関係より多い。これは、多数の意味が混在していることが主な原因である。よって、本手法で扱うには、意味関係の表現方法を整理することが必要である。

6. 一般の複合名詞に対する解析実験と考察

5 章の実験では、情報処理分野の専門用語に対して提案手法を適用し、有効であることが明かになった。今後、この手法の拡張にあたって、情報処理分野で構築した手法がどれだけの見通しを持って、一般の複合名詞に対して適用できるかを現段階で予備的に調べておくことは重要である。そこで、新聞記事に出現する、一般的な分野の複合名詞に対して、提案手法がどの程度有効であるか、解析実験ならびに考察を行う。

6.1 解析対象の複合名詞

1999 年度の日本経済新聞の記事から乱数によって取り出した、2 単語からなる動詞由来複合語 415 語について係り関係の解析実験を行う。主辞のサ変名詞は専門用語の実験の際、構築した辞書に存在するものを選んだ。また、複合名詞として独立して意味をなしているかどうかの判断は人手により行った。

6.2 解析モデルと実験結果

解析モデルは 5 章と同じものとする。ただし、主辞

たとえば「音声認識」は内項と解析されて「音声を認識する」と表現できる。

たとえば「常識推論」は「常識を用いて推論する」といい替えられる。

のサ変名詞は TLCS の辞書があるが、名詞の場合は辞書にない場合、新たに属性を付与する必要がある。主辞のサ変名詞の異なり語が 104 語、名詞の異なり語は 380 語であり、そのうち、新たに名詞の属性を付与したのは 336 語である。提案する解析モデルは 415 語中、412 語 (99.3%) の複合名詞を項関係か修飾関係かを正しく判断した。

6.3 考 察

上記の実験では 415 語中、3 語について解析誤りとなった。誤りの種類は 2 つに分類でき、すでに 5.2 節の考察で観測されたものと同様である。以下、順に説明する。まず、TLCS パターンとの組合せのルールで例外となり、正しくとけなかった複合名詞である。

{	営業補償
	営業損失

上段は項関係で下段は修飾関係である (以下、同じ)。「補償」「損失」は TLCS の 2 と 3 に対応する。これらのサ変名詞の TLCS は 5.2 節と同じく意味述語「NOT」の有無が違ふ。5.2 節の結果もふまえて、一般の複合名詞に対して、「NOT」による係り関係の影響の有無について、調べる必要がある。

残りの 2 語は、名詞の分類と主辞の TLCS のタイプがまったく同じであるにもかかわらず、係り関係が異なったものである。

{	水平調整	{	自己記述
	水平命令		自己記録

各々の主辞のサ変名詞「調整」「命令」「記述」「記録」の TLCS は 2 であり、提案する枠組みではうまくとけない。

上記のタイプの語についてうまく記述できなかったが、そのような語の出現頻度は少ないので、良い性質を持った知識であると考えられる。まだ解析対象とする複合名詞が少ないが、新聞記事に出現するような一般的な複合名詞に対しても、提案手法の見通しは良いといえる。

7. おわりに

本報告では、動詞由来複合語内の係り関係について、分野に依存しない語彙的性質に基づいた、複合名詞解析モデルを構築した。語彙的性質として LCS を用い、LCS をもとにした動詞の分類 (TLCS) と名詞の分類を提案した。これらの分類を行い、TLCS 内の意味述語に対する関係を体系的に整備し、一貫して付与しておくことで、複合名詞の係り関係がとらえられること

を明らかにした。

今回の実験のために構築した、動詞の TLCS 分類と名詞の分類は分野に依存しない言語データであると仮定している。基本的には人手で動詞と名詞を分類することを考えているが、より大きな辞書を構築するには、機械処理によってコーパスを利用して、動詞と名詞の格関係を精度高く獲得できれば、大変有用である。機械処理の可能性について検討していきたい。

評価実験は情報処理用語辞典の 2 単語からなる専門用語 816 語のうち 99.4% の語を正しく、内項関係か修飾関係かに解析した。さらに、新聞記事中の一般的な複合名詞に対しても、415 語のうち、99.3% の語を正しく解析した。この結果から、文脈に依存する解析を行う前の語彙的情報として LCS を選択することは有効であることが分かった。

今後、より広い範囲での解析を進めていくと同時に、修飾関係をさらに分類する解析モデルを構築する予定である。

謝辞 新聞記事を使用させていただいた日本経済新聞社に感謝の意を表します。また、1999 年に神戸で行われた日本言語学会夏期講座において、有益な示唆をいただいた影山先生ならびに、受講生の方々に感謝いたします。また、LSC 勉強会において、有益な示唆をいただいている伊藤たかね先生、島村礼子先生、杉岡洋子先生、ならびに、議論していただいた方々に深く感謝いたします。

参 考 文 献

- 1) Grimshaw, J.: *Argument Structure*, MIT Press (1990).
- 2) Kageura, K., Yoshioka, M. and Koyama, T.: Towards a Common Testbed for Corpus-based Terminology: Lexical Units, POS Information and their Utilisation, *1st International Symposium on Advanced Informatics (AdInfo2000)* (2000).
- 3) Lauer, M.: *Designing Statistical Language Learners: Experiments on Noun Compounds*, Ph.D. Thesis, Department of Computing Macquarie University (1995).
- 4) Tsuji, K. and Kageura, K.: An HMM-based Method for Segmenting Japanese Terms and Keywords based on Domain-Specific Bilingual Corpora, *Proc. 4th Natural Language Processing Pacific Rim Symposium* (1997).
- 5) 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, Vol.31, No.11, pp.1696-1699 (1990).
- 6) 小林義行: コーパスを用いた日本語複合名詞の

解析に関する研究, 技術報告 96TR-0002, 東京工業大学理工学研究科 (1995).

- 7) 太田 悟, 宮崎正弘: 複合語用例データベースを用いた複合名詞の構造的曖昧さの絞り込み法, 情報処理学会第 53 回全国大会 (平成 8 年後期), 2-9 (1996).
- 8) 相磯秀夫: 情報処理用語辞典 (コンパクト版), オーム社 (1993).
- 9) 内山将夫, 板橋秀一: 共起関係を利用した日本語名詞の分割, 情報処理学会自然言語処理研究会, 92-NL-91, pp.47-54 (1992).
- 10) 横山晶一, 佐久間一弘: 意味素性を用いた複合名詞の生成による分析, 計量国語学, Vol.20, No.7, pp.304-314 (1996).
- 11) 飯田 仁, 小倉健太郎, 野村浩郷: 英語複合名詞句構成の意味関係と意味処理, 情報処理学会自然言語処理研究会, 46-4, pp.1-8 (1984).
- 12) 宮崎正弘, 池原 悟, 横尾昭男: 複合語の構造化に基づく対訳辞書の単語結合型辞書引き, 情報処理学会論文誌, Vol.34, No.4, pp.743-752 (1993).
- 13) 松本裕治ほか: 日本語形態素解析システム『茶筌』version 2.2.1 使用説明書 (2000).
- 14) 影山太郎: 文法と語形成, ひつじ書房 (1993).
- 15) 影山太郎: 動詞意味論, くろしお出版 (1996).
- 16) 影山太郎: 形態論と意味, くろしお出版 (1999).
- 17) 竹内孔一, 内山清子, 吉岡真治, 影浦 峽, 小山照夫: 語彙の制約を考慮した複合語解析モデルの構築, 情報処理学会自然言語処理研究会, 2000-NL-136, pp.71-78 (2000).
- 18) 西野哲朗, 藤崎哲之助: 漢字複合語の確率的構造解析, 情報処理学会論文誌, Vol.29, No.11, pp.1034-1044 (1988).
- 19) 内山清子, 竹内孔一, 吉岡真治, 影浦 峽, 小山照夫: 複合名詞の語構成要素間の結合関係に関する考察, 信学技報, TL99-6, pp.7-12 (1999).
- 20) 乾 裕子, 元吉文男, 井佐原均: 語構成論に基づいたサ変動詞の分類, 情報処理学会自然言語処理研究会, 95-NL-110, pp.103-110 (1995).
- 21) 杉岡洋子: 動詞の意味構造と付加詞表現の投射, 研究報告 (2) 先端的言語理論の構築とその多角的な実証 (2-A), pp.341-363, 神田外語大学 (1998).
- 22) 柏野和佳子, 小椋秀樹, 田中牧郎, 加藤安彦: 話し言葉コーパスにおける単位区切りと品詞付与方法, 言語処理学会第 6 回全国大会 (2000).

(平成 13 年 3 月 16 日受付)

(平成 14 年 2 月 13 日採録)



竹内 孔一 (正会員)

1991 年姫路工業大学工学部電子工学科卒業. 1998 年奈良先端科学技術大学院大学博士後期課程修了. 博士 (工学). 同年学術情報センター助手. 2000 年国立情報学研究所助手, 現在に至る. 辞書知識に基づく言語解析モデルの構築, 統計的手法に基づく言語モデルの構築と応用, 専門用語解析のためのコーパス作成等の研究に従事. 電子情報通信学会, 言語処理学会各会員.



内山 清子 (正会員)

1997 年慶應義塾大学大学院政策・メディア研究科修士課程修了後, 1998 年から 2000 年まで学術情報センターの COE 研究員を経て, 現在, 慶應義塾大学大学院政策・メディア研究科博士課程に在籍中. 自然言語処理, 機械翻訳の研究に従事. 言語処理学会学生会員.



吉岡 真治 (正会員)

1991 年東京大学工学部精密機械工学科卒業. 1996 年同大学大学院博士課程修了. 博士 (工学). 同年学術情報センター助手. 2000 年国立情報学研究所助手. 2001 年北海道大学大学院工学研究科助教授, 現在に至る. 知識ベースに基づく設計支援システムの構築, 設計過程のモデル化, 専門用語解析のためのコーパス作成等の研究に従事. 人工知能学会, 日本機械学会, 精密工学会, ASME 各会員.



影浦 峽

1986 年東京大学教育学部卒業. 1988 年同大学大学院教育学研究科修士課程修了. 同年学術情報センター研究開発部助手. 1994 年同助教授. 2001 年より国立情報学研究所助教授. 情報メディア分析, 専門用語研究に従事. 図書館情報学会, 言語処理学会, International Quantitative Linguistic Association 各会員.



小山 照夫(正会員)

1972年東京大学工学部機械工学科卒業．1977年同大学大学院博士課程修了．工学博士．東京都老人総合研究所研究員，浜松医科大学助教授，学術情報センター助教授，学術情報センター助教授/教授を経て2000年国立情報学研究所教授．医療情報処理システム，知識の表現と利用，専門用語解析，用語解析のためのコーパス作成等の研究に従事．電子情報通信学会，人工知能学会各会員．
