

複合名詞解析モデルにおける動詞に対する語彙概念構造の付与法

竹内 孔一

(国立情報学研究所)

1 はじめに

従来、複合名詞内の係り関係の解析手法として、統計的手法に基づくもの [1][2] もの、や、応用問題に特化したルールに基づく手法 [3] が提案されてきた。しかし、統計的手法は学習に使用する事例の分野と質によって解析精度が大きく左右され、また、応用問題に特化した手法は他の問題に対する拡張性が不透明であるため、問題ごとにモデルを見直す必要が出てくる。これに対し、本研究手法では、複合名詞解析モデルを 1) 構成する語に関する語彙情報に基づく解析と、2) 文脈や分野に依存する情報に基づく解析の 2 つに分けることで、分野に柔軟に対応する汎用性のあるモデルの構築を目指す。

現在、1 の観点から、語彙概念構造 (Lexical Conceptual Structure, 以下, LCS [4][5]) を基にした解析手法を提案し、解析実験を行なっている。その結果、複合名詞内の名詞と動詞間の係り関係を解析する手法として、LCS が有効であることを示した [6] が、そこでは、LCS の構築と付与法が重要な課題となった。

そこで、本研究発表では、複合名詞の係り関係を解析する観点から現段階で構築できている LCS パターンとその付与法について議論する。そして、付与した LCS に対して複合名詞解析の観点から考察を加え、有効性について検討する。

2 LCS による複合名詞の解析

2.1 係り関係の整理

形態論における文法的知見 [4][5] から、主辞が動詞の場合、前項の名詞と動詞との関係は大きく項関係か修飾かの 2 つに分けることができる。例えば、

- 機械 翻訳
- 機械 操作

の場合、上側の例では「機械」が主辞の「翻訳」を修飾する解釈であり、下側の例では通常、「機械」を操

作」という内項の解釈となる。係り関係を解析する第一の段階として、項関係か修飾関係かをまず分類することを解析の目標とする。

2.2 LCS による複合名詞の係り関係解析モデル

上例の主辞動詞 (サ変名詞) を LCS 表記で表すと

翻訳 [x CONTROL [BECOME [y BE AT z]]] (1)
操作 x ACT ON y (2)

となる。「x」は動作主、「y」は目的語を表し、「ACT」、「CONTROL」、「ON」、「BECOME」、「BE」、「AT」は基本的な意味述語を表している。「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。一方、内項関係の「機械操作」の場合は、「機械」は (2) 式の「ON」の項の「y」に代入される形となる (下図参照)。

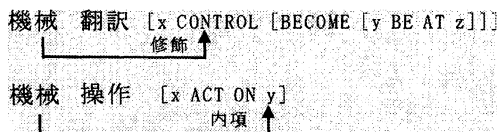


図 1: 「機械」の係り先

そこで解析モデルでは、サ変名詞を中心とした動詞を LCS で分類し、LCS のパターンに対して名詞を分類することで、修飾関係か内項関係を決定する (詳細は文献 [6] を参照)。

3 複合名詞解析モデルのための LCS

本研究で提案する複合名詞解析モデルは内項関係なのか、修飾関係になるかを決定するモデルである。つまり、LCS における内項に対する操作の違いが解析のポイントとなる。よって内項に対する操作を重視して、文献 [4][5] に報告されている LCS を基に、複合名詞解析のための動詞に付与する LCS のパターンを作成した。

全ての動詞に対する LCS パターンがどの程度になるのかは、まだ、明らかになっていないが、本研究では、情報処理用語辞典に記載されているサ変名詞を中心に、新聞記事による事例も含めて約 250 語の動詞を分類している。基本的には 1 つの動詞に対して 1 つの LCS を付与するものと仮定している。

以下では、現段階での LCS パターンを報告し、使用した意味述語について説明する。次に、LCS 分類のための付与法について記述する。

3.1 LCS パターン

- 1 [x ACT ON y] 操作, 運転
- 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] 移動, 遷移
- 8 [x CONTROL[y BE AT z]] 管理, 維持
- 9 [x CONTROL[BECOME[x BE WITH y]]]
認識, 予測
- 10 [y BE AT z] 関係, 位置
- 11 [x ACT] 会議, 行列
- 12 [x CONTROL[BECOME [(NAME OF x) BE AT z]]] 署名
- 13 [x CONTROL[BECOME x BE WITH [(FEELING OF x) BE AT y]]] 尊敬, 信用
- 14 [BECOME[y BE WITH [(FEELING OF y) BE AT z]]] 感動

‘CONTROL’¹, ‘BECOME’, ‘MOVE’, ‘TO’, ‘ACT’, ‘ON’, ‘BE’, ‘AT’, ‘WITH’, ‘NAME’, ‘OF’, ‘NOT’, ‘FEELING’ が意味述語を表し、また、角括弧 ‘[]’ で括られる範囲で句を表しており、一つの意味まとまりをなす。LCS はこれらの句がいくつか組合わさってできている。

¹ 英語では ‘CAUSE’ で記述される。これは日本語の場合、「燃やしたけれど、燃えなかった」のように、必ずしも行動が結果を起こす因果関係になっていない場合があるためである [5]。

‘x’, ‘y’, ‘z’ が各々、主格、対格、与格を与えられた名詞が入る変更 (variable) を示す。‘x’ を外項、‘y’ と ‘z’² を内項と呼ぶ。外項 x は複合の際に目的語を取り込むことは無い (外項排他の制約 [5])。

3.2 LCS パターンの説明

LCS パターンを説明するために、まず、LCS で用いた構成要素について説明を与え、次に、項の取り方、完了性により、横断的に説明する。

- ‘y BE AT z’ は ‘y’ がある状態 ‘z’ に存在することを表している。
- ‘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している。
- ‘y MOVE TO z’ は ‘y’ が ‘z’ という場所に移動することを表している。
- ‘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している。
- ‘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、例えば、「信頼が回復する」「信頼を回復する」のように「れる」などの助動詞を必要とせず他動詞表現と自動詞表現が可能なことをいう。これは能格動詞と呼ばれ、漢語動詞に観測される³。この ‘x=y’ の性質をもつのは LCS 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している。
- ‘x ACT’ は ‘x’ の継続的な動作を表している。
- ‘x BE WITH y’ は ‘x’ が ‘y’ と一体化していることを表している。
- ‘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している。
- ‘[NAME OF]’ や ‘[FEELING OF]’ は通常 ‘y’ である内項の位置が埋まっているため、複合名詞において、内項関係を取らないことを表している。‘NAME OF’ や ‘FEELING OF’ の述語自身に深い意味は無い⁴。

² ‘z’ は与格の項である場合以外に、目的状態として項が埋まっている場合の 2 種あるが、違いは省略している。

³ 英語の他動詞にも観測される。‘break’ など。

⁴ この LCS は複合名詞解析における内項に焦点をあてて構築しているため、より深い意味的な違い (‘尊敬’ と ‘信用’ の意味的な違い) は考慮していない。

3.2.1 完了性

‘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 LCS 番号の 2~7, 9, 12~14 は完了性を持つ。完了性を持つ動詞に対しては、「している」という助動詞を付与してテストを行なうことができる。例えば、LCS 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、LCS 番号 1 と 11 の動詞には完了性はない。例えば、1 の動詞では「操作している」のように今行なっている動作の意味となる。

3.2.2 項の取り方と動詞の分類

‘ACT ON’ か ‘CONTROL’ 項が LCS 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。この場合 ‘x’ はガ格をとらない、内項 ‘y’ はヲ格をとまって出現する (LCS の 1~5, 8, 9, 13 が該当する)。

逆に、‘ACT ON’ か ‘CONTROL’ 述語が無いものが自動詞である。自動詞は、さらに種類があり、‘BECOME’、‘MOVE’、‘BE AT’ がトップにある構造は非対格自動詞 (LCS の 6, 7, 10, 14)、であり、内項はガ格をとまって出現する。

‘ACT’ があるか、もしくは、内項が埋まっている場合は非能格自動詞 (LCS の 11, 12⁵) で、外項はガ格をとまって出現する。能格動詞 (LCS の 5) は自動詞と他動詞の両方の性質を持つ。

3.3 動詞に対する LCS の付与

動詞に対して、上記 14 種類の LCS パターンを付与する。まず、大域的な特徴である、1) 動詞の格の取り方による他動詞、自動詞の分類と、2) 動詞の完了性を利用して、大まかに分類し、次に個別の違いを検証していく。1) 2) に関する特性と格 LCS パターンとの関係を表 1 に示す。

ここで、表中の番号は 3.1 節で示した LCS 番号である。以下では、分類操作について記述する。

1. 動詞の格の取り方を調べる。

- 対象とする動詞の必須格 (ガ, ヲ, ニ格) の取り方を調べる。ヲ格を取る場合、通常、他動詞か能格動詞である。

⁵ ニ格をとるので、不完全非能格自動詞と呼ぶ。

表 1: LCS のカテゴリー分け

動詞の種類	完了性	
	あり	なし
他動詞	2,3,4,13 9	1 8
非対格自動詞	6,7,14	10
非能格自動詞	12	11
能格動詞	5	

- 自他交替を起こすものは、能格動詞として LCS 番号 5 のパターンを付与する。残りが他動詞である⁶。
- ヲ格が無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 LCS は 2~4, 6~7, 9, 12~14 となる。
- 完了の解釈が無い場合、その動詞の LCS は 1, 8, 10, 11 となる。

3. 以下、複数の LCS の部分を分類する。

他動詞で完了性のある動詞 (表 1 中の左上)

- 意味的に分類する。ニ格の名詞を付与して移動の意味がある場合⁷ LCS の 4 に分類。
- ヲ格をとる名詞が変化してのではなく、主体の知覚や考えなどが変化する場合 LCS の 9 か 13 である。ここで、「~に対して~の気持ちを持つ」という言い替えができるもの⁸は LCS の 13 に分類。そうでない場合は LCS の 9 に分類する。
- 残った LCS パターン 2 と 3 は ‘NOT’ の述語の有無の違いである。否定的な動作がある場合は LCS の 3 に分類し、そうでない

⁶ この時、受身形を確認しておく。「ミスを生じる」→「*ミスが生じられる」(*印は非文を表す)では、「生じる」は他動詞ではない [5]。

⁷ 「データを相手に伝送」

⁸ 例えば、「先生に尊敬の気持ちを持つ」。

い場合は 2 に分類する。

4. 他動詞で完了性が無い動詞 (表 1 中の右上)

主体の動作性の意味の強いものが LCS の 1 である。「たくさん～している」を付与して、主体の動作量が多い場合⁹ は LCS の 1 に、結果できたものの量を指す場合¹⁰ は LCS の 8 に分類する。

5. 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)

- 上記でも用いたように、二格の名詞を付与して移動の意味がある場合¹¹ LCS の 7 に分類する。
- 「～に対して～の気持ちをもつ」という言い替えができるもの¹² を LCS の 14 に分類し、残りを 6 に分類する。

3.4 LCS の内項の性質に対する考察

上記で提案した LCS パターンは複合名詞の係り関係において、内項関係かどうかを解析するために導入した。基本的には、LCS 内に内項 'y' があれば、文章中の表現では対格をとる動詞であるが、なぜか、経験的に、内項関係として複合しない語が検出された。以下、LCS パターンごとに例をあげて、説明する。

● LCS 番号 10

分類される動詞として、和語は「ある」、漢語は「位置」、「関係」、「連続」がある。和語の場合は「教師あり学習」における「教師がある」という表現のように、内項をとる複合名詞が存在する。しかし、「停車位置」「師弟関係」のように、漢語の場合は内項とる事例がいまのところ観測されていない。この考察から、この LCS の漢語の場合は、複合化で修飾関係のみという規則を得る。

● LCS 番号 12

「署名」というサ変名詞であるが、通常の対格における内項の部分が「主体の名前」で埋まっている。よって「女性の名前を署名する」ならば、ヲ格をとることができるが、「*名前署名」と複合化しない。その変わり、与格の項がある

ため、「条約署名」(「条約に署名」) 与格に対して項関係になることができる。

● LCS 番号 13 と 14

「尊敬」や「感動」であるが、これらの動詞は、表層的にはヲ格、二格をとることができるが、複合名詞としては現れない。(例えば、「先生を尊敬」→「*先生尊敬」、「音楽に感動」→「*音楽感動」。) これらも、内項に対する意味構造が通常の動詞とは異なることに起因しており、それが LCS パターンで表現できている。

4 おわりに

本報告では、複合名詞の係り関係の解析のための語彙概念構造 (LCS) の提示と付与法を明らかにした。まだ、記述の第一歩であり、今後、より広い範囲で解析を進めていくことで、修正と拡張を進めていきたい。

5 謝辞

新聞記事を使用させていただいた日経新聞社に感謝の意を表します。

参考文献

- [1] M. Lauer. *Designing Statistical Language Learners: Experiments on Noun Compounds*. PhD thesis, Department of Computing Macquarie University, 1995.
- [2] 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. Technical Report 96TR-0002, 東京工業大学理工学研究科, 1995.
- [3] 宮崎 正弘, 池原 悟, and 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, 34(4):743-752, 1993.
- [4] 影山 太郎. 文法と語形成. ひつじ書房, 1993.
- [5] 影山 太郎. 動詞意味論. くろしお出版, 1996.
- [6] 竹内孔一, 内山清子, 吉岡真治, 影浦 峯, and 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. In 情報処理学会自然言語処理研究会, 2000-NL-136, pages 71-78, 2000.

⁹ 「たくさん操作している」は操作量が多い。

¹⁰ 「たくさん管理している」は管理しているものが多い。

¹¹ 「地面に移動する」

¹² 例えば、「建築に感動の気持ちを持つ」