

複合名詞語構成要素のための品詞相当カテゴリーに関する予備的考察

内山 清子 竹内 孔一 吉岡 真治 影浦 峽 小山 照夫
学術情報センター

1. はじめに

情報検索や機械翻訳において、専門用語の扱いは一括して辞書に登録する方法が主流であるが、日々新しい用語が生成されるため、辞書登録では対応しきれない。そこで複合語を構成する語構成要素の特徴から複合語を生成するメカニズムが明らかになれば、一語として辞書に登録されていなくても解析が可能になる。そのためにも複合語の語構成要素の特徴を、詳細に整理・分析していくことが必要とされる。

2. 研究の背景

2.1 従来の文／単語関係における品詞について

従来は、複合語の生成や解析について、複合語を構成している個々の要素を眺めて複合語内で起こり得る結合のパターンや制約、意味関係に関する研究 [1] [2] が中心であった。既存の複合語について、語形成の結合を解析する方法は効率的であり、規則を抽出しやすい。しかし結合の意味関係を中心とした研究が多く、基本的な形態素、構文などの形態的な情報に関する解析があまり明確になっていない。形態面からどこまで解析が可能であるかを調べることにより、今まで行われてきた意味中心の解析手法と合わせて、より精度の高い複合語解析が行えると考えられる。

また、従来の文／単語関係で定義された品詞を複合語の語構成要素にそのまま適用することは難しい。たとえば、「学際的研究」の「学際」は従来の品詞では名詞と定義されるが、同じ名詞である「人間」「機械」のように「が」格「を」格を後続し主語となり得る語とは異なり、「的」などの接尾辞と接続して語を修飾する役割を果たす。同様に複合語の語構成要素には名詞と同形の形容動詞語幹も含まれている。「安全保障」という複合語の語構成要素は「安全」と「保障」であるが、「安全」は「安全を保障する」のように「を」格を後続して目的語となることができるため、表面的には名詞と分類できるが、「安全運転」では「安全な運転」という形容動詞語幹となっている語もある。このように、名詞に属する要素一つ取ってもその用法は複合語中ではさまざまである。本研究では、専門分野における複合語の語構成要素に対応可能な品詞（以下品詞相当カテゴリーと呼ぶ）を作成し定義づけるための予備的検討を試みる。

2.2 従来の品詞定義について

従来の文／単語関係で定義された品詞のうち、名詞に分類されるカテゴリーを対象にして、複合名詞の解析に適した品詞相当カテゴリーの設定について研究を進めた。まず、従来の品詞定義の内容を把握するために参考としたのは、岩波の国語辞典 [3] と EDR 電子化辞書 [4] の品詞体系である。岩波による名詞の定義では、①呼び名として使い、②格助詞がつく、③活用をしない、④多くのものは主語として使えるが、そうでないものも名詞に含まれる場合がある」としている。EDR 電子化辞書における普通名詞は、「格助詞『が、を』が後続するもの」と定義している。そこで複合名詞の品詞相当カテゴリーと定義を設定するとき、上記の 2 つの定義で対応が可能であるか、またどのような問題があるのかを考えた。

2.3 従来の名詞定義における問題点

上記の名詞定義のうち主語となり得る可能性と格助詞接続について検討した。多くのものは主語となり得るという定義について、「が」格「を」格を後続するパターンは、単純に格助詞の接続だけでは扱えないのではないかという問題があがった。それは単独で「が」「を」を後続する場合と、「〇〇の」という語が前接して結合した複合形式で主語となる場合があるからである。たとえば意識は「意識がなくなる」「意識をとりもどす」というように単独で「が」「を」を後続するが、上昇は「温度の上昇が起こる」「相場の上昇を支えた」は「温度の」「相場の」を含めた複合の形式で「が」「を」を後続して主語となり得る。また、「が」「を」を後続しないが、その他の格助詞「の」「に」「へ」「で」などを後続する語（主語にはなり得ないもの）の可能性も考えられる。さらに、前述した通り従来の品詞相当カテゴリーの名詞に分類される語の中に形容動詞的なふるまいをする語が潜在的に存在するなどの問題も含まれている。これらの問題を考慮して複合名詞の解析に適した品詞相当カテゴリーの検討を行った。

3. 品詞相当カテゴリーの設定と定義の検証

3.1 品詞相当カテゴリーと定義の設定について

従来の名詞に分類していたカテゴリーにおける問題点を整理して、名詞性、形容動詞性、名詞性と形容動詞性、その他の品詞相当カテゴリーの 4 つのカテゴリーを

設定し、それぞれについて詳細な定義を考えた。

(1) 名詞性品詞相当カテゴリの定義

・単独

N1: 「が」「を」が後続して主語になり得るもの

N2: 「が」「を」を取らず主語にはなり得ないが、
「の」「に」「で」「へ」「から」「と」「で」「より」の
格助詞が後続するもの

N3: 「の」だけが後続するもの

・複合

N4: 指示語が前に接続して「が」「を」が後続して主語と
なり得るもの

N5: 「A の B」のように「A の」が前に接続して「が」「を」
が後続して主語となり得るもの

(2) 形容動詞性品詞相当カテゴリの定義

A1: ダ活用形容動詞語幹「ナ活用」

清らか (だ), きれい (だ)

A2: ダ活用形容動詞語幹+連体「の」あり

「ナノ活用」 無心 (の)

A3: タルト活用形容動詞語幹

堂々 (たる), 洋々 (たる)

(3) 名詞+形容動詞性品詞カテゴリの定義

名詞性品詞相当カテゴリと形容動詞性品詞カテゴリ
の両方の定義に当てはまる場合、+印でつなげて記
述する。例: N1 と A1 の要素を持つ場合 N1+A1

(3) その他カテゴリの定義

E: 上記の名詞あるいは形容動詞性品詞カテゴリの定義
に当てはまらないもの

3.2 分類作業

上記の品詞相当カテゴリを検証するために、この
定義に沿って分類を行った。分類対象とする語基は、情
報処理用語辞典 [5] から抽出した複合語を、学術用語語
基表 [6] の語基の定義に基づいて分割したものである。
分割して得られた語基から2語漢語1803語を取り出し、
ランダムに500語を選んだ。作業は対象語基が仮説のど
の品詞相当カテゴリにあてはまるかを作業分担者が個
人の主観により判断した。N2, N3 については N4, N5 に
も該当するかどうかを調べ、該当する場合は+印をつけ
て記述する。そして、作業結果を1人のデータだけで検
証することは難しいことや、ある程度一般的な傾向を測
るために分類作業は4人で行った。

3.3 分類結果

分類した結果では、4人全員が一致した品詞が少なか
ったため、3人以上一致した数を表1に示した。N1 が圧
倒的に一致した数が多く、ついで A1, A2, E の順で一致

数が多い。それ以外の品詞カテゴリは一致度が低く、
個人間でばらつきが見られた。N1 の「が」「を」を後続
して主語となり得るものが数的に一番多いことも一因で
あるが、主観的な判断としては一番結果が収束しやすい
基準であることが明らかになった。

表1: 被験者3人が一致した品詞カテゴリと一致数

品詞カテゴリ	一致数	品詞カテゴリ	一致数
N1	226	N2+A2	4
A1	37	N1+A2	2
A2	14	N2+A1	2
E	14	N3	2
N1+A1	8	N2+N5	1
N2	4	N3+A1	1

3.4 分類結果の考察

分類結果について特に N2~N5 の名詞性品詞カテゴリ
の判定が個人間でばらついていた。原因は、設定した
品詞相当カテゴリの定義のうち特に名詞に関する部分
が適切でない、あるいは、手続き、作業上の問題という
2つの点が考えられる。手続き、作業上の問題点に関し
ては、品詞相当カテゴリが適切な設定であるという前
提で調べることが可能であるが、品詞カテゴリそのも
のが確定できない現段階では難しい。そこで手続き、作
業に関しては問題がないと仮定し、品詞相当カテゴリ
の設定に関して問題点を明確にしていく。

3.5 新聞記事の用例収集と分類

品詞相当カテゴリの定義の妥当性を評価するため
に、個人の主観によるばらつきの原因を解明しなければ
ならない。そこで個人の主観とは独立した客観的で一般
的なデータと比較することにした。データは日本経済新
聞94年度版のCD-ROM版を利用し、そこから用例を抽出
する作業を試みた。個人の主観による分類作業で使用し
た語構成要素を新聞記事データから検索し、N1 から E
までの仮説の品詞相当カテゴリに当てはまる用例を手
作業で抽出した。

3.5.1 個人主観と新聞記事用例の結果の比較検討

日本経済新聞のデータから得られた結果を一つの指標
として、個人の主観による判定結果を比較して、仮説の
品詞相当カテゴリの定義について検討を行った。その
際以下の点に分けて比較した。

1. 被験者の3人以上が一致したカテゴリと新聞記事デ
ータの用例との比較
2. 被験者の判定結果が一致しなかったものと新聞記事デ
ータの用例との比較

まず1については、作業分担者 3 人以上が一致した個人主観による判定結果は、作業人数としては数が少ないが、人間にある程度共通する情報であると考えられる。そのため新聞記事データの結果との間に差がでない語構成要素については、仮説の品詞相当カテゴリーの設定が妥当なものであることが証明できる。しかし、結果が異なる語構成要素は人間の主観と新聞記事データによる用例との間に何らかの言語現象の差異が生じていることになる。

次に2については、被験者間でそれぞればらつく傾向が見られる語構成要素が、実際の新聞の用例ではどのように使用されているのかを調べ、ばらつき現象の原因を明らかにすることにより、品詞相当カテゴリーの設定が適切であったかどうかを検討していく。

3.5.2 個人主観が3人以上一致した判定結果について

表2に個人主観の判定が3人以上一致した品詞相当カテゴリーと新聞記事データによる判定結果を比較した内容を示す。(ここでは被験者3人がN1で一致した語構成要素を除いている)

表2：個人主観の判定が3人以上一致したカテゴリーと新聞記事データとの比較結果

比較結果	該当数
一致しない	40
A1で一致	20
Eで一致	10
A2で一致	7
新聞記事データがない	5
N2で一致	3
N1A1で一致	2
N1+A2で一致	1
N3で一致	1
合計	89

3人以上一致した個人主観の結果と新聞記事データによる結果を比較してみると、比較的A1, E, A2においては一致が見られたが、N2からN5はほとんど一致していない。また、N4, A3の定義に当てはまる語がなかったため、N4, A3の定義は不必要であることが明らかになった。次に一致していない部分を調べてみると、いくつかのパターンを見ることができた。

(1) 個人主観ではA1に一致していたものが、新聞記事ではN1の用例もあった。

複雑：「戦いは複雑を極める」

確実：「今回は確実を期す」

正確：「足取りは正確を極めた」

(2) 個人主観ではA2に一致していたが、新聞記事ではナ活用が検索できなかったため、N2となったもの。

自明：「自明の前提」

正規：「正規の用紙」

(3) 個人主観ではE（文中においては単独で使われず、「的」などの接尾辞を伴うもの）と判定されたが、N2であったもの。

能動：「女性の意識が能動に転じた」

国際：「国際の平和を願う」

間接：「間接の経済波及効果」

(1) について通常「複雑を」「確実を」というようにに格助詞「を」を直接伴って使用するとは考えにくい。「が」「を」を後続する多くの場合は、「複雑さを痛感した」「複雑さが増す」や「正確さが損なわれる」「正確さを確保する」など、接尾辞「さ」を伴っている。直接「が」「を」を後続している用例を眺めてみると、必ず「極める」「期す」という動詞を使っていることから、動詞との共起関係が影響しているものと考えられる。

(2) は、新聞記事データが1年分しかないということもあり、データが不十分であったことも理由にあげられるが、「自明な」「正規な」という活用は考えられても実際の文で使用するかどうかは疑問がある。今後データを増やして調べていく必要がある。

(3) は、「国際政治」「間接照明」など、複合形式で使うことが多い語であるが、「の」「に」といった格助詞を直接後続した用例が検索された。これは「国際平和」が「国際の平和」と直接言い換えることができるわけではないが、複合語から転じてこのような表現が可能になったのか、あるいは文表現における助詞の省略形として複合語に発展したのかどうかは、今後興味のある問題である。

このように個人の主観と新聞記事の用例による判定結果の一致率は半分しかなかったため、仮説の品詞相当カテゴリーの妥当性を評価するまでには至らなかった。しかし、一致しなかった部分における言語現象の違いを明らかにできた。

3.5.3 個人主観の判定が一致しないカテゴリーについて

個人主観の判定結果が個人間で揺れていたために品詞カテゴリーを決定できなかった語構成要素について調査を行った。新聞記事データの用例から格助詞などの接続情報を抽出し、その結果に基づいて、品詞カテゴリーを暫定的に決定した。表3にその結果を示す。

表3：新聞記事データの用例による判定で決定された
主な品詞カテゴリー一覧

品詞カテゴリー	該当数	品詞カテゴリー	該当数
N2	38	N3	9
N1	35	E	7
N2+N5	26	A2	6
N5	12	その他	37

個人主観が作業分担者の間で揺れが生じてばらついていたものは、新聞記事データの用例からN2と判定されたものが一番多く、次にN1という結果になった。つまりN1,N2,N5の判定が個人間で揺れているパターンが多かった。

この現象は、単独で主語となるのか、複合形式で主語となり得るかを判断する時に個人間の主観に差があったからだと考えられる。そして、そのような判断の揺れが大きい語は、新聞記事の用例においては「が」「を」を後続して主語になっていることが多い。その原因は、新聞記事の場合、あるトピックに関する流れの中で語を使用するために、文脈の中における語の使用例が抽出できるので、語を中心に考えた例よりも使用範囲が広がっているのではないかということである。そしてこれは人によって判定時に想定した文脈という制約の範囲が異なるため差が生じたのではないだろうか。

品詞相当カテゴリーの定義設定の段階において、単独で主語になる語と複合形式「～の」を伴った状態で主語になり得るものという2つのパターンがあることは既にわかっていて、「～の」という複合形式の接続パターンも一種の文脈による制約と言える。さらに新聞記事データから抽出した用例では、単純に「～の」がついて「が」「を」を後続するパターンだけでなく、さまざまな形式の文脈による制約が得られた。個人の主観で揺れていた原因がこうした文脈による制約の違いにあるのではないか。そこで文脈による制約にはどのような種類があるのかを調べた。

4. 文脈による制約について

新聞記事データで N1、N5 に判定された語構成要素の用例を集め、その中から文脈による制約を受けている語について、以下の3つの例をあげる。

(1) 前の文脈を必ず受ける語

「両立」について、「両立が難しい」など単独で使用することは可能であるが、実際は「学業とスポーツの両立が難しい」など、「両立」には2つの事象を指す特徴がある。

(2) 形容動詞などの修飾語が前述する語

「学術」は、「学術が培われた」のように単なる「学術が」ではなく「多彩な」という修飾語がついて「多彩な学術が培われた」という形式で使用される傾向がある。

(3) 省略形として使われる語

「学会参加費は一般が1000円である」の場合の「一般」は学会員ではない一般の人のことを指している。つまりこの文脈において「一般の人」を省略して「一般」という表現をしている。

このような文脈による制約について、さらに詳細な分析を行って整理し、その結果を品詞相当カテゴリーの定義に反映させていく方法を今後の課題として検討していく。

5. まとめ

従来の文/単語関係による品詞カテゴリーでは複合語解析で不都合が生じる部分を明らかにするために、複合語解析に有効な品詞相当カテゴリーを従来の名詞に属するカテゴリーを対象にして仮説を提案し、分類を行った。分類は個人の主観による判定と新聞記事データから抽出した用例による判定の2つを行い、2種類の判定による差の原因について考察し、同時に仮説の品詞相当カテゴリーの検証を行った。さらに検証の過程で、個人主観の差は語を使用する文脈制約の違いにあるという可能性を指摘し、今後の研究課題とした。

参考文献

- [1] 野村雅昭，四字漢語の構造，国立国語研究所報告 54，電子計算機による国語研究Ⅶ，秀英出版，1974.
- [2] 野村雅昭，石井正彦，学術用語の造語法、特定研究『情報化社会における言語の標準化』成果報告，1986.
- [3] 相磯秀夫，情報処理用語大辞典，コンパクト版，オーム社，1993.
- [4] 野村雅昭，石井正彦，学術用語語基表，1989.
- [5] 西尾実，岩淵悦太郎，水谷静夫，岩波国語辞典，第5版，岩波書店，1997.
- [6] EDR 電子化辞書研究所，EDR 電子化辞書使用説明書 version1.5，1995.