

複合名詞の構成要素間の関係推定の一手法

小林義行 徳永健伸 田中穂積
東京工業大学 情報理工学研究所

{yashi,take,tanaka}@cs.titech.ac.jp

1 はじめに

複合名詞は、2つ以上の単純な名詞が結合して構成される名詞である。ここで単純な名詞とは国語辞典の見出し語として記述されるような名詞のことである。単純な名詞から複合名詞は限りなく生成されるので、その全てを辞書に記述することはできない。単純な名詞がどのように結合して複合名詞を構成しているか解析することが必要である。

我々はこれまでに、複合名詞を構成する要素のあいだの係り受けを解析する方法を提案している。例えば、「教育制度改革」の正しい係り受け構造((教育, 制度), 改革)を得ることで、正しく“reformation of the educational system”と翻訳できる。「教育制度改革」を構成する各名詞を翻訳し同じ順にならべた“the education system reformation”は適切な翻訳ではない。

しかし、構成要素のあいだの係り受け構造を解析するのみでは十分ではない。例えば、「自然保護運動」は“nature conservation movement”と訳され、「自然食品店」は“natural food store”と訳される。「自然」は、前者では“nature”、後者では“natural”と訳し分けられている。「自然」を“nature”と訳すためには、「自然」が「保護」の目的語であること、“natural”と訳すためには「自然」が「食品」を連体修飾しているという関係が必要である。このように複合名詞の構成要素のあいだの結合関係を推定しなくてはならない。

言語学や日本語学において、複合名詞を構成する要素のあいだの関係を分類した研究がいくつかなされている。本論文では、上記の複合名詞の構成要素のあいだの関係を推定するという問題を、あらかじめ設定された結合関係の分類のどれに該当するかを推定する問題と考える。そこで、本論文では、1) 計算機による複合名詞の解析に利用できる分類、2) その分類を利用するために必要な言語的知識、3) 言語的知識の獲得方法、について検討する。

2 複合名詞の構造に関する分類

複合名詞を構成する全ての名詞のあいだの結合関係は、係り受け関係にある2つの名詞のあいだの結合関係が推定できれば決定できる。例えば、「所信表明演説」を構成する「所信」「表明」「演説」のあいだの結合関係は、「所信表明」と「演説」のあいだの結合関係と、「所信」と「表明」のあいだの結合関係から決定することができる。そのため、複合名詞の構造を分類する研究では、2つの名詞が結合した複合名詞の結合関係を分類している。

複合名詞を構成する2つの名詞のあいだの結合関係を分析した研究は以下の2つに分けることができる[5]。

1. 構造論的分類 構成要素を品詞性、語種などに基づいて分類し、構成要素の組み合わせによって複合名詞を分類したもの。基本的には、名詞を、1) 格になる名詞、2) 動作を表す名詞、3) 形容詞的な名詞に分類し、その組み合わせによって複合名詞を分類する。
2. 意論的分類 連体修飾句へのパラフレーズに基づいて、要素のあいだの関係を、例えば「並立、主述、補足、修飾、補助、客体」のように分類したもの。

2つの分類方法は全く無関係ではなく、文の統語構造と意味構造のような関係になっている。例えば、複合名詞「自然保護」は、構造論的には「自然(格になる名詞)」+「保護(動作を表す名詞)」という構造で、意味論的には「自然(目的語)+保護(動作)」という補足の構造である。この関係は「自然を保護する」という文において、統語構造が「自然を(後置詞句)+保護する(述語動詞句)」で意味構造が「保護(object=自然)」であるという関係に対応している。

自然言語処理では、複合名詞の構造論的な分類を求め、その結果に基づいて意味論的な分類を求めるといった処理の流れが自然であると思われる。

構造的な分類としては、野村の分類を利用する[4]。野村はまず名詞を統語的な性質に基づいて、以下のように分類している。

体言類 N	格助詞「ガ・ヲ・ニ…」をともなう
用言類 V	スルをともない述語成分になる
相言類 A	ナ(ノ)をともない体言類を修飾する
副言類 M	ニをともない用言を修飾する
接辞 F	接頭辞と接尾辞

この分類を以下では名詞の品詞性分類と呼ぶ。

これらの名詞の品詞性分類の組み合わせによって、複合名詞には以下の構造が考えられる。

表 1 構造論的分類

構造名	組み合わせ	例
補足	N+A N+V	栄養+豊富 原価+計算
	V+A V+N	実現+可能 有+意義
連用修飾	A+V V+V	特別+参加 除行+運転
	M+V	直接+交渉
連体修飾	A+N V+N	高級+官僚 勤労+意欲
	N+N	化学+肥料
並列	N-N A-A	党利+党略 自由+自在
	V-V	自給+自足
対立	N/N A/A	竜頭+蛇尾 西高+東低
	V/V	男尊+女卑
重複	X=X	近々
補助	F+X X+F	不+明瞭 弱体+化
省略		経済 農協
省借		葡萄 砂利

複合名詞の多くは、補足、連用修飾、連体修飾、補助の4つである。本論文における複合名詞の解析では、自立語のみを含む補足、連用修飾、連体修飾の3つの構造を対象にする。

上記の複合名詞の構造分類を推定するためには、名詞の品詞性分類に関する知識が必要である。本研究では、野村の分類を参考に名詞の品詞性分類を以下の3種類に整理し直す。

体言名詞	格要素になりうる名詞。体言類に該当
形容名詞	ナ(ニ)がつくことで、いわゆる形容動詞になりうる名詞。相言類に該当
動名詞	スルがつくことができるいわゆるサ変名詞用言類に該当

単独の用法では副言類である語は、複合名詞の要素となる場合は、相言類とみられるので、副言類に該当する分類は削除する。

名詞がどの分類に該当するかを獲得する知識源としては機械可読辞書を用いる。機械可読辞書には、サ変名詞になりうるか、形容動詞になりうるかについての記述がなされている。この情報を用いて、動名詞に分類される語、形容名詞に分類される語を求めることができる。

3 構成要素の関係の推定

複合名詞の構造分類を同定するのみでなく、構成要素がどのような関係であるのか推定することが必要である。例えば、構造分類を同定することで、「自然保護」の構成が[自然(体言名詞)+保護(動名詞):補足]と分かるが、[自然(体言名詞)+ヲ(格助詞)+保護(動名詞):補足]というように自然が保護のヲ格要素であることを推定することが望まれる。

複合名詞の構造論的分類と意味を関係付けると以下の表になる。意味は名詞句へのパラフレーズによって表現する。

表 2 構造と意味の関係

品詞性分類の組	構造名	意味
体言名詞+動名詞	補足	政治ヲ改革スル
体言名詞+形容名詞		在庫ガ過剰ダ
動名詞+形容名詞		実現(スルコト)ガ可能ダ
形容名詞+動名詞	連用修飾	新規ニ開店スル
動名詞+動名詞		除行シテ運転スル
形容名詞+体言名詞	連体修飾	異常ナ事態
動名詞+体言名詞		影響スル因子
体言名詞+体言名詞		宇宙ニアル基地
		外科ガ専門デアル病院

複合名詞の構成が補足関係の場合、構成要素のあいだに補充される格助詞を推定するできれば、2つ要素のあいだの意味的關係を求めるのに利用でいける。このとき、述語要素となっている名詞がとりうる格の情報が必要になる。このために、コーパスや機械可読辞書の例文から得られる[体言名詞+格助詞+動名詞/形容名詞]という共起を抽出して利用することが考えられる。また、動名詞を他動詞、非対格自動詞、非能格自動詞に分類することは、格助詞を推定する場合の手助けとなる。例えば、他動詞、非能格自動詞の主語は複合名詞の要素にならないという分析が利用できる[3]。

連体修飾、連用修飾のあいだに成り立つ意味關係の一部も格助詞を補間することで推定できる。[形容名詞/動名詞 A 体言名詞 B]という構造に対して[体言名詞 B 格助詞 形容名詞/動名詞 A]という共起データがある場合である。例えば、「演算装置」に対して「装置デ演算」と

いう共起データがある場合「デ」は道具格なので「演算」からみれば「装置」は道具とわかり、「装置」からみれば「演算」はその用途を意味しているとわかる。

体言名詞と体言名詞の関係は、補充できる述語を用いて推定することが考えられる。名詞句「名詞₁+ノ+名詞₂」に対して名詞₁と名詞₂の意味的關係を推定する知識を、名詞₁あるいは名詞₂を格要素としてとりうる動詞や形容詞の情報を用いて決定しようとする研究がある[6]。この研究では、IPALで動詞、形容詞の意味を分類するのに用いる述語素によって、2つの名詞の意味的關係を推定している。例えば、IPAL形容詞辞書に「海が色が青い」という例文があるとき、この場合の「青い」の述語素はPVである。述語素PVより「色」は「青」の属性であると分かる。この方法は、複合名詞における2つの体言名詞の意味的關係の推定に利用できると考えられる。

4 分析結果

複合名詞を分析して、各構造論的類別の分布を調査した。また、既存の機械可読な知識を用いてどの程度、意味的な解析が可能か実験した。

2つの名詞からなる複合名詞としては、田中が収集した4文字漢字語[7]のうち分類語彙表[2]に記述されている2文字語2語に分割できる121151語を用いた。

構造の推定

機械可読辞書から獲得した品詞性分類に関する知識を用いて複合名詞の構造分類を推定した。品詞性分類は大辞林から獲得し、サ変動詞になると記述されている名詞を動名詞に、形容動詞になると記述されている名詞を形容名詞に、それ以外の名詞を体言名詞に分類した。複合名詞を構成する2つの名詞それぞれの品詞性分類を求め、その組合せによって構造論的類別を推定した。例えば、「安全+運転」は「安全」は形容名詞で、「運転」は動名詞なので構造は「形容名詞+動名詞」と推定する。ここで、すべての名詞は1つの品詞性分類にのみ該当すると仮定している。結果は表3に示す。各構造に該当すると解析した複合名詞の数をデータ数の列に示す。正解率は、推定した構造の正しさを無作為抽出で100語を選び検査した。

格助詞の補間

構造分類が補足関係に該当する複合名詞に対して格助詞を補間する実験を行なった。知識源としては、新聞記事から「名詞句+格助詞+述部」という関係を抽出した田中のデータを用いた[8]。格助詞はガラニデの4種類である。格助詞補間の方法は以下の通りである。

1. 知識源 [名詞句+格助詞+述部] から述部の主要素が動名詞あるいは形容名詞であるものを抽出
2. 抽出した知識源 [名詞句+格助詞+動名詞/形容名詞] の名詞句から主辞を抽出し、解析用知識 [名詞+格助詞+動名詞/形容名詞] を生成
3. 複合名詞 [名詞 A+名詞 B] に対して解析用知識に [名詞 A+格助詞 J+動名詞/形容名詞 B] があれば助詞 J を補間

表3 構造の分布

構造	データ数	正解率 [%]
補足関係		
体言名詞+動名詞	27289	87
体言名詞+形容名詞	1195	57
動名詞+形容名詞	1308	85
連用修飾		
動名詞+動名詞	18692	47
形容名詞+動名詞	3985	68
連体修飾		
体言名詞+体言名詞	29209	99
動名詞+体言名詞	32241	82
形容名詞+体言名詞	5153	73
その他		
形容名詞+形容名詞	382	22
合計	119454	

さらに、知識源 [名詞 A+格助詞+動名詞/形容名詞 B] の名詞、動名詞、形容名詞を分類語彙表での分類(以下意味分類)に置き換え、複合名詞 [名詞 A+名詞 B] の名詞を意味分類に置き換えたものに対して上記の方法で格助詞の補間ができるかを実験した。

結果を表4に示す。行c+cは複合名詞を構成する名詞を両方とも意味分類に置き換えたとき、行w+cは右の名詞を意味分類に置き換えたとき、行c+wは左の名詞を意味分類に置き換えたとき、行w+wは意味分類を用いないときの格助詞を補間できた複合名詞の数である。データ数は格助詞を補間した複合名詞の数である。正解率は、補間した格助詞が正しさを構造ごとに無作為抽出で100語を抽出して検査した。ただし、100語以下しか該当する結果がない構造については全ての結果を検査した。複数の解を出力した場合は、不正解とした。

自動詞・他動詞の推定

大辞林[1]の例文を用いて動名詞が自動詞的か他動詞的か推定する実験を行なった。無作為に選択した他動詞10語、非能格自動詞10語、非対格自動詞10語の例文の特徴を分析した結果、自動詞の例文ではガ格を含むも

のあるいはヲ格を含まないものが多く、他動詞の例文ではヲ格を含む場合は他動詞が多いことが分かった。例文にヲ格以外の格助詞を含む場合は自動詞的、ヲ格を含む場合は他動詞的として、動名詞が自動詞的か他動詞的か推定し、その知識複合名詞の解析に用いた。

表 4 解析結果 (括弧内は%)

構造	c+c		w+c		c+w		w+w	
	データ数	正解率	データ数	正解率	データ数	正解率	データ数	正解率
補足関係								
体言名詞+動名詞	20691	25	5097	57	7091	60	1480	77
体言名詞+形容名詞	299	17	49	34	1	0	33	97
動名詞+形容名詞	278	10	16	68	4	25	6	83

大辞林に記述されている動名詞は 13357 語で、自動詞的と判定されたものは 2576 語、他動詞的と判定されたものは 2708 語であった。この知識を格助詞 (ガまたはヲ) の補間を用いた結果、補足構造 [体言名詞+動名詞] 27289 語に対して 16522 語に格助詞を補間できた、正解率は 55% であった。

5 考察

品詞性分類の情報のみでは十分な正解率で構造を推定することができなかった。その原因の一つは、野村の分類に当てはまらない複合名詞が存在することにある。[形容名詞+形容名詞] とならぶ構造は、野村の分類では並立構造しかない。これに該当するのは例えば「自由自在」である。しかし、「十分可能」のように連用修飾構造のものがあることが分かった。また、[動名詞+動名詞] は連用修飾関係のみであるが補足関係になる場合もある。「気化+冷却」は連用修飾関係であるが、「改革推進」は補足関係である。

もう一つの原因は、本論文で名詞が複数の品詞性分類に該当する場合を考慮していないことによる。例えば、「安全」は、体言名詞と形容名詞の両方に該当する。「安全運転」では「安全」は形容名詞で、「安全な運転」あるいは「安全に運転する」のように連体修飾語あるいは連用修飾語として用いられている。一方、「安全確保」では体言名詞で「安全を確保する」のように格要素になっている。

格要素になりうるかどうかは機械可読辞書には記述されていない。そのため、ある名詞が動名詞あるいは形容名詞になるかどうかは分かるが、体言名詞になりうるかどうかは分からない。つまり、動名詞あるいは形容名詞でしかないのか、体言名詞でもあるのか辞書のみでは分からない。

「安全」は大辞林では名詞と形容動詞と記述されてい

るが、同じように名詞と形容動詞になりうると記述されている「豊富」は「安全」と同じようには格要素にならない。「豊富」の場合、格要素になりうるのは派生語「豊富さ」である。動名詞や形容名詞に該当する名詞が体言名詞にも該当するかを知るには、コーパスを利用して実

際の用法から情報を得る必要がある。

コーパスから得られる共起情報を用いて格助詞の補間を試みたが、この共起情報を用いることで、構造分類の正解率も同時に向上させることができる。ただし、構成要素をシソーラスの意味分類に置き換えると正解率が低下する。

6 おわりに

本論文では、名詞を統語的な性質に基づいて分類し、その名詞の分類に基づいて複合名詞を構成する要素のあいだの関係を推定する方法について述べた。この方法は、機械可読な知識源から得られる情報のみを用いているので、人手による規則の記述に伴う開発の困難さという問題を軽減することができると思われる。

今後の課題は、考察で述べたような問題を解決して実際の解析に運用できるように、解析に用いる情報を整備することである。また、連体修飾関係、連用修飾関係に対する構成要素間に利用できる知識についての考察も残されている。

参考文献

- [1] 松村明 (編) . 大辞林. 三省堂, 1988.
- [2] 林大. 分類語彙表. 秀英出版, 1966.
- [3] 影山太郎. 文法と語構成. ひつじ書房, 1993.
- [4] 野村雅昭. 二字漢字語の構造. 日本語学, Vol. 7, No. 5, 5 1988.
- [5] 沖久雄. 複合名詞の意味と構文. 日本語学, Vol. 2, No. 12, 12 1983.
- [6] 奥村学, 仲記弘. 形容詞・動詞辞書を用いた名詞間の意味的關係の決定. IPAL シンポジウム'95 論文集, 1995.
- [7] 田中康仁. 自然言語の知識獲得-四文字漢字列. 情報処理学会第 45 回全国大会, 1992.
- [8] 田中康仁. 語と語の關係解析用データが, を, に, でを中心にして