

3C-2

複合名詞における 語と語の関係について

石崎雅人

NTT情報通信処理研究所

1. はじめに

複合名詞は「文の凝縮」と言われるように〔1〕その中に含まれる語と語には多様な関係が存在する。例えば機械翻訳システムにおいてよりよい訳語を得ようとする場合にはこの関係を決定する必要がある。複合名詞の中の語と語の関係を決定するための情報として利用できるものとしては品詞情報、浅い意味情報（概念階層における位置）、深い意味情報（概念の構造）が存在する。関係を正確に決定するためには深い意味情報を考慮する方が望ましいが、このような構造を辞書中のすべての語彙に対して無矛盾に構築するのは難しい。ここではより一般性のある浅い意味情報を用いた関係の決定について考察する。浅い意味情報は個々の語に対して複数個付与される。概念階層をネットワークと考えると、語と語との関係を決定することは多義に相当する概念間の複数のリンクから正しい関係を表現しているリンクを選択することであると考えることができる。このように考えると語と語の関係を決定する際に問題となるのは以下の2点である。

- ①どのようにして複数のリンクから最も確からしいリンクを選択するか。
- ②リンクがどのような関係をあらわしているか。

本研究では①の問題を中心に考察する。①の問題のために概念階層におけるリンクに対して条件を付与し、それを使って最も確からしいリンクを選択する方法を提案する。2章では複合名詞の中の関係を定義する。3章では複合名詞の語と語の関係を決定するために必要な条件について述べる。4章では3章で述べた条件付きの概念階層をもとにした関係の決定方法及びその決定率を調べた実験について考察する。

2. 複合名詞の中の関係の定義

日本語の複合名詞の中の語と語の関係についてはいろいろな関係が提案されている〔1〕、〔2〕、〔6〕。これらは品詞と格関係を中心にしたものが多く十分に整理されているとはいいがたい。また助詞『の』で結ばれる名詞句の関係〔4〕については十分に整理されてはいるが、必ずしもすべて複合名詞に適用できるとは限らない。そこで複合名詞の中の語と語の関係を整理した。これら

は『並列』、『同格』、『修飾』、『格修飾』にまとめられた。これを中心にしたデータ約2300語の複合名詞に対して関係を調査して図1に示す関係を詳細に定義した〔7〕。

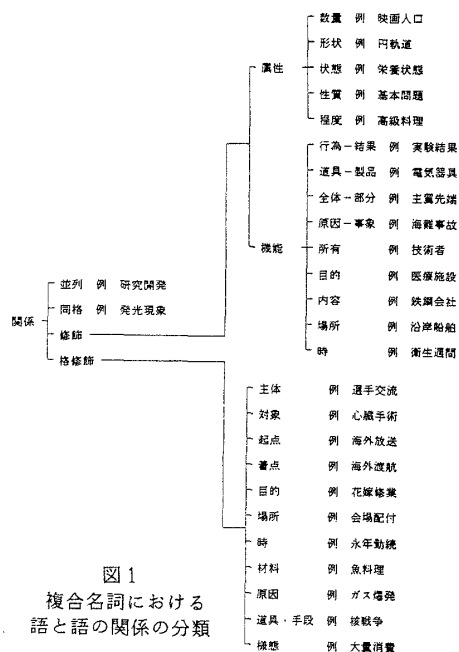


図1
複合名詞における
語と語の関係の分類

3. 関係決定のための条件

3.1. 関係

『関係』は、そのリンクがどのような関係をあらわしているかを示すために用いる。この『関係』の項に記述できるものは図1に示す関係及びNILである。関係がNILの場合、リンクの表わす関係は〔ノード名-ノード名〕であるとする。

3.2. 用言分類

複合名詞の中の語には、用言と同じような働きを持つ語がある。『用言分類』とはそれらを用言とみなして、その格要素の持ち方により分類したものを言う。これは3.3.で述べる『評価値』とともに複数のリンクの中から最も確からしいリンクを選択するために使用される。現時点では用言と同じような働きをする1語が1分類であるとしている。

3.3. 評価値

『評価値』は、複数のリンクの確からしさを経験的な評価関数によって計算したものである。これは用言分類によってリンクが選択できなかったときに最も確からしいリンクを選択するために用いる。この関数はノードの距離に関する関数と頻度関数から成る。頻度関数では試

行データから求めたリンクが正しく張られている概念の対をデータベースを基にノードとノードの共起確率を計算する。このようにして作成した評価関数 $F(cat1, cat2)$ を式(1)に示す。ただしA、Cは重み定数、D、Bはそれぞれ距離の関数、共起確立、cat1、cat2は概念を示す。

$$F(cat1, cat2) = A * D(cat1, cat2) + C * B(cat1, cat2) \quad (1)$$

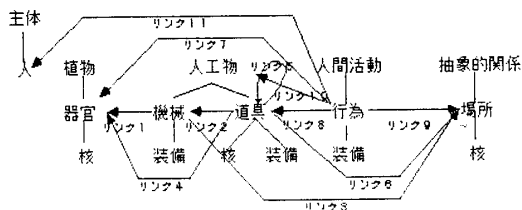
4. 条件付き概念階層を用いた関係の決定

4. 1. 決定方法

関係の選択方法は条件を使用する順序によって規定される。その順序は用言分類→評価値となる。リンクを選択するときにまず用言分類と語が一致するリンクを調べる。もしリンクが複数存在すれば他の条件を利用する。リンクが存在して唯一であればそのリンクが関係を示す。この場合リンクは必ずしも語を直接含む親概念間に張られているわけではないので、概念階層を探索する必要がある場合もある。唯一でなければさらに評価値を用いて、最も点数の高いリンクを選択する。この評価値は相対値であり値そのものに意味があるわけではないので、最高点からある範囲にあるリンクを正しいリンクであるとして選択する。

この決定方法を例語『核装備』の関係の選択に適用する。図2にそのネットワークを示す。関係決定の手順は以下のようなになる。

- (1) 用言分類と語との一致を見る。この場合一致するのでリンク10及びリンク11を候補に残す。
- (2) リンク11が示す概念は他方の語を含まずリンク10は含むのでそれが『核』と『装備』の関係を示す。以上からその関係は「対象」と決定される。



リンクの種類	関係	用言分類	評価値
リンク1	NIL	NIL	410
リンク2	NIL	NIL	470
リンク3	NIL	NIL	470
リンク4	NIL	NIL	410
リンク5	NIL	NIL	551
リンク6	NIL	NIL	470
リンク7	NIL	NIL	410
リンク8	NIL	NIL	532
リンク9	NIL	NIL	480
リンク10	対象	装備	440
リンク11	主体	装備	380

図2 例語『核装備』に関するネットワーク

4. 2. 実験結果

関係の決定方法について約100語の複合名詞を用い

て実験を行なった。評価値を計算するために予め別の200語の複合名詞を基にデータベースを作成しておいた。その実験結果を表1に示す。関係を完全に正しく決定した比率は62.8%であり、1つでも正しく決定されたものを入れると87.6%となった。この方法を用いない場合の決定率rはリンク数をnとすると式(2)によって表わされる。

$$r = 1 / n \quad (2)$$

ノードとノードを結ぶ平均リンク数は4.6であるため、式(2)よりこの方法を用いない場合の決定率は21.7%であり、図3に示すようにこの方法は有効であることがわかる。

正しく決定できた	61
正しく決定できたが、正しくないものまで含んだ	14
正しく決定できたが、すべてではなかった	11
誤って決定した	12
総数	98

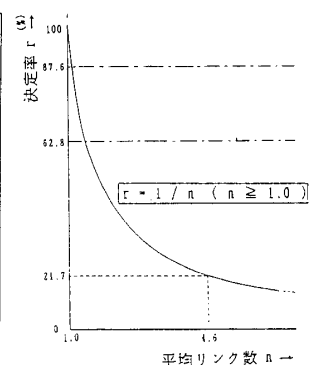


表1 モデルを用いた決定方法に関する実験結果

図3 平均リンク数と関係の決定率の関係

5. おわりに

本論文では以下のことについて述べた。第1に複合名詞の中の語と語の関係を『並列』、『同格』、『修飾』、『格修飾』の4種類に分類した。さらにそれぞれについて約2300語の複合名詞について調査し、詳細に分類した。第2に複合名詞の中の語と語の関係を決定するために、概念階層に条件付けを導入した。第3に条件付き概念階層における語と語の関係を決定する方法について考察した。概念階層をネットワークと考えると、語と語の関係を決定することは、ネットワーク上のノード間の複数のリンクから最適なリンクを選択することであることを述べ、その決定方法について述べた。この方法について実験を行った結果、87.6%という決定率が得られた。この方法を用いない場合の決定率が21.6%であることより、この方法は有効であると考えられる。

6. 参考文献

[1]奥津:「複合名詞の生成文法」国語学101号pp.19-33.
 [2]藤田他:「漢字連続複合語の解析」第28回情処全国大会7M-3 pp.1287-1288. [3]:飯田他:「英語名詞句構成の意味関係と意味処理」NL研46-4(1984). [4]島津他:「日本語文意味構造の分類—名詞句構成を中心に—」NL研47-4(1985). [5]島津他:「助詞「の」が結ぶ名詞の意味関係のsub-categorization」NL研53-1(1986). [6]野村:「四字漢語の構造」国研報告54 pp.36-80(1975). [7]国研:「電子計算機による新聞の語彙調査(IV)」(1973). [8] Isabelle, P.: "Another Look at Nominal Compounds" COLING-84 pp.509-516(1984). [9]:Brachman, R, J. et al.: "An Overview of the KL-ONE Knowledge Representation System" COGNITIVE SCIENCE 9 pp. 217-260(1985).