ショートノート

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†

石崎 雅人††

本論文では，2名詞からなる漢字複合名詞における名詞の意味の多義の解消のためのアルゴリズムを提案した．多義解消のためのデータとして，意味情報・選択制限情報・統計情報を利用している．アルゴリズムは，統計情報がある一定の点数を持つものについて，一方の名詞の選択制限情報と他方の名詞の意味情報との満足度および統計情報の点数により順位付ける．満足度は，複合名詞においてどの格によって名詞と名詞が関係付けられやすいか，および名詞の性質（選択制限情報を持つかどうか）の組合せに関するヒューリスティクスを用いて計算する．このアルゴリズムを98語の複合名詞に適用したところ，87.6%の割合で正しい意味を選択した．

1. はじめに

日本語の技術文章において漢字からなる複合名詞はよく用いられている．複合名詞を含んだ文を解析する場合，これを1単語として辞書に登録すると，(1)辞書の規模が大きくなる，(2)複合名詞間の関係がわからない*，という問題がある．

複合名詞における関係の決定についての研究によると[3)-5)]，(1)名詞間には種々の意味的関係が存在すること，(2)完全に正しく関係を決定するためには世界知識が必要であるが，意味と用言性名詞**が持つ選択制限は関係の決定に有効である，ということがわかる．

名詞は通常複数の選択制限および意味を持っている．複合名詞内の名詞間の関係を決定する際に，複数の可能性の中で正しい選択制限または意味が決定されていれば，関係の決定はより容易になる．

本論文では，意味情報，選択制限情報，統計情報を利用した2名詞からなる漢字複合名詞における名詞の意味の多義の解消アルゴリズムを提案する．以下，2，3章ではそれぞれ，意味の多義に関して利用するデータ構造およびアルゴリズムについて述べる．4章では本アルゴリズムの実験結果について考察する．

2. 多義解消のためのデータ構造

意味の多義解消のために，意味情報，用言性名詞から抽出した選択制限情報および意味情報間の関連の強さを示す統計情報をデータとして保持する（図1参照）．

意味情報は名詞を最上位ノードとし，約70のノードを持つ概念階層（図2参照）の中に位置付けられる．

用言性名詞に関連する用言の選択制限情報は，名詞の字面から予測できる用言が辞書にある場合はその選択制限情報を取り出し，そうでない場合は意味的に関連する用言*** の選択制限情報を取り出す．両方の場合ともデータがない場合は「なし」とする．

複合名詞内の名詞の意味情報の間には結び付きやすさがあると思われる．これを表現するために，統計情報として複合名詞のテストデータから計算した意味情報の共起確率と意味情報間の距離を利用する．

共起確率 stas (sem 1, sem 2) は，ある意味情報の組合せの数 c(sem 1, sem 2) と全体の意味情報の組合せの数 tc との比を計算することで求める((1)式参照)．

(1) $\mathrm{stas}(\mathrm{sem\ 1, sem\ 2}) = c(\mathrm{sem\ 1, sem\ 2})/tc$

「の」で結ばれる名詞句の解析において概念階層体系における意味情報間の距離が重要な役割を持つ[2)]．名詞句と複合名詞には共通点が多いことから，意味間の距離の影響を計算する関数 dist(sem 1, sem 2)（図3参照）に重み定数 a をかけ，共起確率に加えて統計

† An Algorithm for Disambiguating Noun Meanings in Noun-noun Compounds of Kanji Characters by MASATO ISHIZAKI (NTT Communications and Information Processing Laboratories).

†† NTT 情報通信処理研究所

* 関係の決定により，例えばキーワード自動抽出における適合率を上げることができる．

** 用言のように他の名詞と格関係をとる名詞．

*** 意味的に関連する動詞を持つ名詞を Isabelle は role-nominal と呼んでいる．例えば，cat food における food は eat と意味的に関連付けることができる．

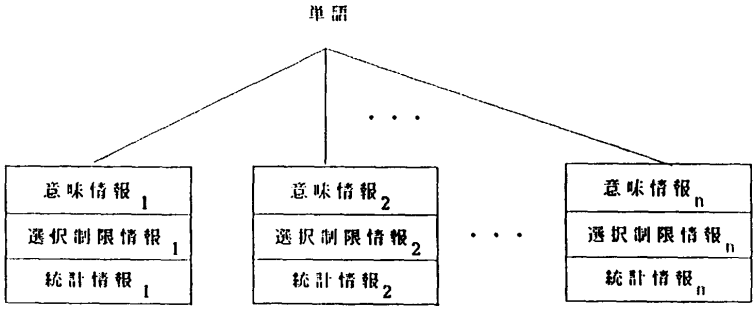

図 1 複合名詞内の名詞の意味の多義を解消するためのデータ構造
Fig. 1 The data structure for disambiguating noun meanings in compounds of kanji character.

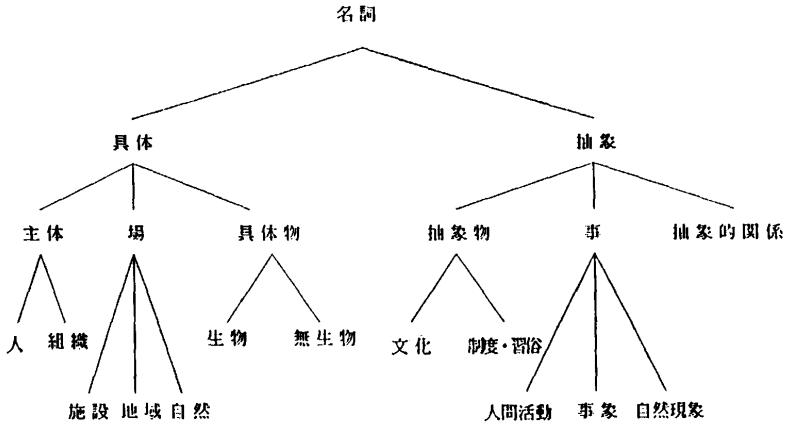

図 2 本論文で使用した概念階層体系（一部）
Fig. 2 A part of the concept hierachical system in this paper.

意味 sem 1 と意味 sem 2 の距離	dist (sem 1, sem 2)
0	500
1	470
2	440
3	410
4	380
5	350

意味 sem 1 と意味 sem 2 の距離は，（1）sem 1 と sem 2 が同一のとき，0，（2）概念階層体系中の高さ i の部分木に sem 1 と sem 2 が含まれるとき，i とする．

図 3 意味の距離の影響を計算する関数 dist の定義
Fig. 3 The definition of the function 'dist' that calculates the effect of semantic distance.

情報 mix (sem 1, sem 2) を求める（(2)式参照）．

(2) mix (sem 1, sem 2)
= stas (sem 1, sem 2)+a*dist(sem 1, sem 2)

「装備」に関するデータを**図 4** に示した．選択制限情報は（(格の名前　格が取り得る意味情報（意味情報　点数))…）という形式を持っている．最後の要素である（意味情報　点数）は選択制限情報の点数付けの結果を保持する．統計情報が未定なのは計算量を減らすため，多義解消時に値を計算するからである．

3. 多義解消のためのアルゴリズム

名詞 N 1, N 2 の意味の多義がそれぞれ m, n 個あるとした場合の多義解消のアルゴリズムを**図 5** に示す．

名詞 N 1 の選択制限情報 N 1. sr が存在するならば，名詞 N 2 の意味情報 N 2. sem が N 1. sr を満足するかどうかを調べる．意味情報が選択制限情報を満足するとは，概念階層体系において意味情報で示されるノードが選択制限情報が示すノードに等しいかまたは選択制限情報のノードが支配する木構造に含まれることをいう．複数の選択制限情報が存在する場合には，用言性名詞がどの格で他の名詞と関係しやすいかを表す格の順序* の高いもので，選択制限情報を他方の名詞の意味情報が最もよく満足するものを選択す

* 目的格>主格>その他の格．

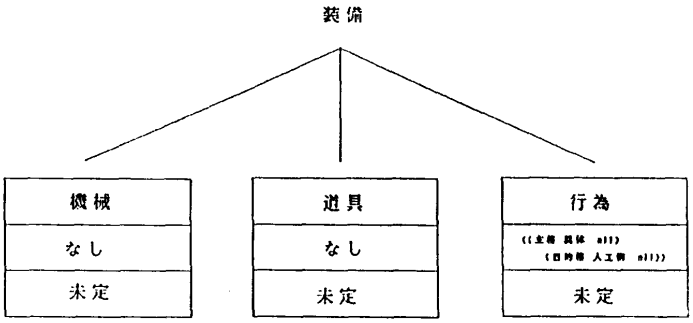

図 4 名詞「装備」に関する多義解消のためのデータ
Fig. 4 The data for disambiguating noun meaning of 'soobi.'

定義：複合名詞は名詞N1、名詞N2から構成されているとする．
意味情報をN.sem、選択制限情報をN.sr、統計情報をN.statと名付ける．
名詞Nにおいて複数のデータがある場合には、$N.sem_i$ $N.sr_i$ ・$N.stat_i$ のようにそれぞれの情報に添字を付ける．
N1、N2はそれぞれm、n個の意味情報があるとする．

アルゴリズム：

```
do  i = 1, m
 if (N1.sr_i に値がある) then
   do  j = 1, n
    N2.sem_j がN1.sr_i の格を満足するかどうかを
    調べ、満足するならば点数付けをする（注）
    N1.stat_j =mix(N1.sem_i ,N2.sem_j )
   end  do
end  do
do  j = 1, n
 if (N2.sr_j に値がある) then
   do  i = 1, m
    N1.sem_i がN2.sr_j の格を満足するかどうかを
    調べ、満足するならば点数付けをする（注）
   end  do
end  do
```

統計情報が500点以上のデータを取り出し、後ろに選択制限情報を持つもの、前に選択制限情報を持つもの、選択制限情報を持たないものと順位付け、さらにそれぞれの分類の中で選択制限情報を持つものを満足度によって順位付ける

（注）格による点数（目的格500点、主格250点、その他の格0点）と意味間の距離による点数を加えて計算する．

図 5 複合名詞内の名詞の意味の多義を解消するためのアルゴリズム
Fig. 5 The algorithm for disambiguating noun meaning in compounds of kanji characters.

る．意味情報が最もよく選択制限情報を満足するとは，意味情報間のノードの距離が最も短いことをいう．選択制限情報の有無に関わらず，名詞 N1 と名詞 N2 との統計情報を関数 mix により計算し，名詞 N1 および N2 の統計情報 N1. stat, N2. stat に代入する．名詞 N2 についても同様に計算を行うが，統計情報は既に求められているので計算は行わない．

複合名詞において用言性名詞がある場合，選択制限情報で示される関係に成りやすいことおよび用言性名詞は後ろにくるパターンが多いことから，統計情報がある一定の点数を持つものを*，用言性名詞を後ろに持つ複合名詞，用言性名詞を前に持つ複合名詞，用言性名詞を持たない複合名詞のように順位付ける．さらに用言性名詞を持つものについては，選択制限情報と意味情報との満足度**，用言性名詞を持たないものについては，統計情報を用いて順位付けを行う．

複合名詞「核装備」内の名詞の意味の多義の解消に，本アルゴリズムを適用した結果できるデータを図 6 に示す．片方向の矢印（←）は，矢印の根元に選択制限情報を持つ名詞があることを示し，両方向の矢印（↔）は両方の名詞に選択制限情報がないことを示している．

まず名詞「核」について，選択制限情報が存在しないので，「核」の意味である器官，道具，場所と「装備」の意味である機械，道具，行為との統計情報を関数 mix によって計算する．例えば器官と機械に関する統計情報は 410 点となっている．次に名詞「装備」について，選択制限情報が存在するので，格の順序に従って選択制限情報と意味情報との満足度を計算する．一定の点数を 500 点としているので**，複合名詞「核装備」内の「核」と「装備」の意味はそれぞれ（1）道具と行為，（2）道具と道具と順位付けられる．

4. 実験と考察

統計情報を計算するため予め複合名詞 200 語を基にデータベースを作成しておき，別の複合名詞 98 語について実験を行った．実験結果によると（図 7 参照），ありえそうな意味の組合せをすべて選択したものは全体の 62.8% であり，ありえそうな意味の組合せを 1 つでも選択したものを入れると 87.6% となった（複合名詞内の名詞の意味の多義の数は平均 4.6 であった）．

決定が失敗した原因としては，（1）統計情報を計算するための複合名詞のデータベースが十分でなかったこと，（2）複合名詞内の名詞の意味の間の共起関係を

* 頻度の低い意味の組合せを持つ複合名詞を排除するために統計情報の下限（本論文では 500 点）を設けた．
** まず格の順序で順位付け，次に意味情報と選択制限情報との距離で順位付ける．

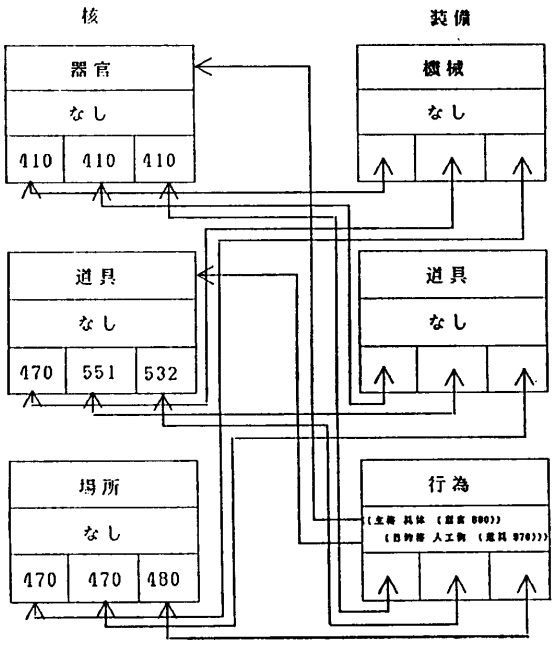

図 6 アルゴリズムを複合名詞「核装備」に適用して得られたデータ
Fig. 6 The data obtained by applying the algorithm to the compound of 'kaku-soobi.'

ありえそうな意味の組合せをすべて選択した.	61
ありえそうな意味の組合せをすべて選択したが, ありえそうにない組合せも選択してしまった.	14
ありえそうな意味の組合せを全部は選択できなかった（ありえそうな意味の組合せの幾つかをありえそうにないとしてしまった）.	11
ありえそうな意味の組合せを1つも選択することができなかった.	12
総　数	98

図 7 複合名詞内の名詞の多義解消アルゴリズムに関する実験結果
Fig. 7 The experimental result of the algorithm.

とらえるのに本論文で使用した概念階層体系が十分に細かくないことが考えられる.

5. 今後の課題

今後は, (1)複合名詞の実験データの充実, (2)応用に応じた意味の詳細化, を行うことによるアルゴリズムの有効性の確認, および本アルゴリズムの複合名詞の解析処理への組込みについて研究していくつもりである. また, 3語以上の複合語への本アルゴリズムの適用については石崎[1]を参照されたい.

謝辞 本研究をすすめるうえで, 有益な御意見を頂いた NTT 情報通信処理研究所寺島信義前自然言語処理部部長, 坂間保雄主幹研究員, 東田正信主幹研究員および同研究部の職員の方々, ならびに ATR 自動翻訳電話研究所森本逞データ処理研究室室長, NTT 基礎研究所島津明主幹研究員に感謝いたします.

参 考 文 献

1) 石崎雅人：日本語複合名詞の解析, 第35回情報処理学会全国大会論文集, 1T-1 (1987).
2) 田村直良, 田中穂積：意味解析に基づく並列名詞句の構造解析, 日本ソフトウェア科学会第3回大会, A-2-1 (1986).
3) Finin, T. W.: *The Semantic Interpretation of Nominal Compounds*, pp. 310-312, Coordinated Science Laboratory, University of Illinois (1980).
4) McDonald, D. and Hayes-Roth, F.: Inferential Searches of Knowledge Network as an Approach to Extensible Language-Understanding Systems, in Waterman, D. A. and Hayes-Roth, F. (eds.), *Pattern-directed Inference Systems*, pp. 431-453, Academic Press (1978).
5) Isabelle, P.: Another Look at Nominal Compounds, *COLING-84*, pp. 509-516 (1984).

(平成2年3月2日受付)
(平成2年9月11日採録)

石崎 雅人（正会員）
昭和35年生. 昭和58年慶應義塾大学電気工学科卒業. 昭和60年同大学大学院理工学研究科電気工学専攻修士課程修了. 同年, 日本電信電話(株)横須賀電気通信研究所入所. 機械翻訳, 自然言語生成に関する研究に従事. 現在, 情報通信処理研究所メッセージシステム研究部研究主任. AIUEO, 人工知能学会, 日本認知科学会各会員.

論 文 誌 編 集 委 員 会

委員長	益田 隆司		
副委員長	名取 亮		
委員	石畑 清	魚田 勝臣	浮田 輝彦
	大田 友一	小池 誠彦	小谷 善行
	佐藤 興二	島津 明	戸川 隼人
	徳田 雄洋	永田 守男	原田 紀夫
	松田 晃一	三浦 孝夫	毛利 友治
	吉澤 康文		