2R−2

# 名詞句「NP の NP」の意味構造推定のための名詞の統語範疇の獲得

田中 省作　冨浦 洋一　日高 達
(九州大学大学院 システム情報科学研究科)

## 1 はじめに

日本語文中には, 2 つの名詞句が助詞「の」で結合した名詞句「*NP の NP*」が頻繁に現れる. この名詞句「*NP の NP*」は, 表層的には単純であるが意味は多様で, 自然言語処理の重要な研究テーマの一つとなっている. このような名詞句「*NP の NP*」に対して, Montague の形式化に従って統語構造と意味構造を対応させた文法体系 [2] が提案されている. この文法体系では, 従来, 単一の統語範疇として扱われていた名詞句を意味的観点から 4 つの統語範疇に細分化し, 統語構造と意味構造との関係を対応づけられている.

本研究では, このような文法体系 [2] に基づいて名詞句「*NP の NP*」の意味構造推定を計算機上で行うために必要となる名詞の統語範疇を, コーパスより獲得する手法を示す.

## 2 名詞句の統語範疇の細分化

[2] で提案されている名詞句「*NP の NP*」の文法体系では, 名詞句を 4 つの統語範疇, 普通名詞句 (CN), 項句 (T), 関係名詞句 (RN), 事象名詞句 (EN) に細分化する. 以下に, 簡単にこれらの統語範疇について説明する.

- 普通名詞句 (common noun phrase; CN)
  性質を表しているような名詞句で, 例えば「色白」「美人」などが CN で, それぞれ "色白である", "美人である" という性質を意味している. また, 統語的には, 「ある」「その」というような限量子が結合することができる.

- 項句 (term; T)
  ある特定の個体や事象を指示しているような名詞句で, 「太郎」や「その人」というような固有名詞や CN に限量子が付加した名詞句などが T である.

- 関係名詞句 (relation noun phrase; RN)
  個体や事象間の関係を表す名詞句で, 「兄」や「目的」などが RN である. 例えば, 「兄」という名詞句は, T の「太郎」を「の」で連接して「太郎の兄」という名詞句を構成し, "太郎の兄である人" を指示することになる. また, このときの T を RN 「兄」の補項, 「太郎の兄」で指示する対象を主項とよぶ.

- 事象名詞句 (event noun phrase; EN)
  動詞性の事象を指示する名詞句で, 「努力」や「考え」というものがある. 動詞が名詞化したような名詞は全てこの範疇に属す.

名詞句「*NP の NP*」の意味構造を正しく推定するには, 少なくともまず単語レベルの任意の名詞が上記の 4 つの範疇のいづれに属すかを正しく決定しなければならない. 未登録語である場合を除いて, 全ての名詞を国文法での文法情報を参照することで, T,EN について全てを網羅できる. 代名詞や固有名詞はある特定の個体を指示しているので T に, サ変動詞の語幹や動詞の連用形と同じ表現になっている名詞は EN に属し, また, 形容詞の語幹に「さ」や「み」などが付いた表現となっている名詞は RN に属すと決定できる. 実際に, 九州大学の公用データベース日本語単語辞書に登録されている名詞 61270 語について調べてみると, 22290 語が T,EN,RN のいづれかに分類される. しかし, 残りの約 63.6%の名詞については, CN,RN のいづれに属すのかは, 品詞や形態などの文法情報からでは判定できない.

## 3 普通名詞句と関係名詞句の統計的性質

コーパス中の名詞句「$NP_1$ の $NP_2$」の大量の用例から, $NP_1$ の意味範疇の散らばりに着目し, CN,RN の統計的性質について検討する.

名詞句「$NP_1$ の $NP_2$」において, $NP_2$ の名詞句によって, $NP_1$ に出現し得る名詞句を多少なりとも限定する. 例えば, $NP_2$ が「机」の場合, $NP_1$ には「木製」「研究室」などは出現可能で, $NP_1$,$NP_2$ の間にはなんらかの意味的な関連性が存在しなければならず, $NP_2$ によっ

---

The Acquisition of Syntax Categories of Nouns for the Estimation of Semantic Structure of Japanese Noun Phrases "*NP 'no' NP*", Shosaku TANAKA, Yoichi TOMIURA, Toru HITAKA, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-81, JAPAN

て $NP_1$ に出現しうる名詞句の意味範疇を限定している．名詞句「$NP_1$ の $NP_2$」において，$NP_2$ の名詞句が，CN か RN によって $NP_1$ の意味範疇がどのように限定されるか，以下に説明する．

### $NP_2$ が CN の場合

1. 「CN の CN」
この場合，「黒茶毛の犬」「秋田犬の犬」「大型の犬」などが挙げられる．このとき，$NP_1$ の「犬」に対して修飾可能なものであれば $NP_1$ に出現し得ることができる．よって，$NP_1$ の意味範疇の散らばりは大きい．

2. 「T の CN」
この場合は，$NP_1$ と $NP_2$ の２つの名詞句の間に，表現としては現れていないある意味関係が成立する．例えば，「太郎の車」という場合，『所有関係』が成立していると考えられる．このとき，$NP_1$ には「車」の所有者になれるような《人》や《会社》などが出現可能である．しかし，$NP_2$ が同じ「車」であっても，$NP_1$ が「駐車場」というような場合，導出される意味関係は『位置関係』となる．この場合，$NP_1$ には場所を表す《地名》や《場所》などが出現可能である．このように，「T の CN」の場合，$NP_1$ に出現し得る名詞句の意味範疇は，$NP_1$,$NP_2$ 間に仮定される意味関係によって，限定される．しかし，仮定される意味関係は，一般に複数存在すると考えられ，その分，意味範疇への限定作用は緩くなり，意味範疇の散らばりも大きくなる．

### $NP_2$ が RN の場合

1. 「CN の RN」
この場合は，「色白の母」や「黒い髪の母」など，RN の主項が，$NP_1$ の CN によって限定される．つまり，$NP_1$ には，$NP_2$ の主項を修飾可能なものであれば，出現し得ることができる．よって，$NP_1$ の意味範疇の散らばりは大きい．

2. 「T の RN」
例えば，「私の妹」「太郎の母」「そのプロジェクトの目的」などがある．この場合は $NP_2$ の主辞の関係名詞の補項に $NP_1$ が割り当てられる．「母」であれば一般に，ある人物とその母親の写像を表しているので，「の」に前置される $NP_1$ は「私」や「母」というような《人》に強く限定される．よって，意味範疇の散らばりは小さくなる．

よって，意味範疇の散らばりは，RN に比べ CN の方が意味範疇が散らばる傾向があることが予想される．

## 4　意味範疇の散らばりの定量化

この名詞句の意味範疇の散らばりを，名詞 $N_1, N_2$ が名詞句「$N_1$ の $N_2$」という形での共起情報を基に，エントロピーを用いて定量化を行う．意味範疇の集合を $C = \{c_1, c_2, \cdots, c_m\}$ としたとき，名詞 $n$ の意味範疇の散らばりを，

$$\mathcal{H}_C(n) = -\sum_{c \in C} P_r(c|n) \log P_r(c|n) \qquad (1)$$

とする．ただし，

$$P_r(c|n) = \frac{\sum_{n';c \text{ の下位語}} f(\langle n', n\rangle)}{\sum_{c \in C} \sum_{n';c \text{ の下位語}} f(\langle n', n\rangle)}$$

また，$f(\langle n', n\rangle)$ は，コーパス中で「$n'$ の $n$」という形で出現した頻度を表す．$n_{\mathrm{CN}}, n_{\mathrm{RN}}$ をそれぞれ CN,RN に属す名詞とすると，$n_{\mathrm{CN}}, n_{\mathrm{RN}}$ が $N_2$ となっているような共起情報が十分えられている場合には，前のセクションで説明したように，CN の方が意味範疇の散らばりが大きくなる傾向があるので，式 (1) では，

$$\mathcal{H}_C(n_{\mathrm{CN}}) > \mathcal{H}_C(n_{\mathrm{RN}})$$

となりやすい．

そこで，式 (1) の値 $\mathcal{H}_C(n)$ を名詞 $n$ の特徴量として，パターン認識手法を用いた判別実験を行った．サンプルは，EDR 日本語コーパスおよび RWC コーパス中の名詞句「$NP_1$ の $N_2$」から $\langle$[$NP_1$ の主辞],$N_2\rangle$ という共起情報を抽出し，その中から $\langle *, n\rangle$ という形で 100 回以上出現したものをデータとした．また，意味範疇の集合として分類語彙表中の項目を使用した．このとき，正解率が 80.9%で，CN,RN が判別された．

## 5　おわりに

本研究では，名詞句「$NP$ の $NP$」の意味構造推定に必須となる名詞の統語範疇の獲得手法について示した．また，従来，明確とされていなかった CN,RN の基準について一つの客観的基準を与えた．現在，名詞句の表層表現中には現れないが $NP_1$,$NP_2$ 間に仮定される意味関係の獲得および推定法について検討している．

## 参考文献

[1] 田中 省作; 名詞句「$NP$ の $NP$」の意味構造推定のための知識獲得, 九州大学 修士論文, 1997

[2] 冨浦 洋一, 中村 貞吾, 日高 達; 名詞句「$NP$ の $NP$」の意味構造, 情報処理学会論文誌, Vol.36, No.6, 1995