

名詞の統語・意味的制約と用例を利用した 日本語名詞句構造解析法

2C-1

江尻 秀彰

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

日本語文には、いくつかの名詞を「の」や「と」で結合した名詞句が数多く出現し、その意味も多様である。このような名詞句は、単純であるがゆえに曖昧さが伴うことが多い。それ故、日本語の処理において、このような名詞句を解析することは、重要で困難な課題となっている。

本稿では、「の」で結ばれた名詞句に限定し、名詞の接続強度（名詞の統語・意味的制約）の利用と、「名詞+の+名詞」型名詞句の用例を利用した、ハイブリッド型の名詞句構造解析を提案し、その有効性を示す。

2 接続強度について

「の」で連結される名詞の場合、後方の名詞が主名詞となり、前方の名詞がこれを修飾することが多い。このため、名詞の性質によって前方に来やすい名詞、後方に来やすい名詞がある。この性質を考慮して、形態素辞書 [1] をもとに名詞を 14 種に分類したものに接続強度 [2] づけした。なお、接続強度は、1 から 5 の 5 段階で記述され数字が大きいほど接続強度は大きく、接続しやすいということを表している。

Japanese Noun phrase Structure Analysis using Syntax, Semantic Constraint and Noun phrase Corpus

Hideaki Ejiri, Masahiro Miyazaki
Niigata University

表 1. 接続強度の例

	左側接続強度	右側接続強度
具体名詞	4	3
抽象名詞	3	4
関係名詞	3	3
サ変名詞	2	1

3 名詞句用例データベース

名詞句解析のため名詞句の用例を集めたデータベースを用意する。ここで、名詞句の用例としては、「名詞+の+名詞」型のものを用意する。これらは、主に EDR 共起辞書から獲得したものであり、約 6 万例を用意した。名詞句の情報としては、各名詞の字面と品詞コードを記述した。

4 類似度について

名詞句用例データベースから用例を検索する際に、その類似度を設定しておくことが大切である。そこで用例数が多いことを考慮して、一致条件を厳しくし、どちらか一方の名詞の字面が一致することを最低条件として、その類似度を設定した。ここでは、主名詞となりやすい後方の名詞と一致する場合に高い点を与えた。

表 2. 一致条件とその類似度

一致条件	類似度
字面 + 字面	1.0
品詞 + 字面	0.8
字面 + 品詞	0.5

5 評価値の計算

入力名詞句の各名詞には、品詞コードから左右の接続強度が与えられ、2つの名詞間の接続強度は、その和である。また、用例データベースとのマッチングをとることにより、その2名詞間の類似度が求められる。ここで、各評価には重みを与える。この重みのとり方でどちらの評価を重視するかが決まる。最終的な結合度は、次式に示すように接続強度と類似度に重みを与えたものの和で表される。

$$C_{total} = W_c \cdot C_{connect} + W_s \cdot C_{similar}$$

C_{total} … 式の評価値
 $C_{connect}$ … 接続度による評価
 $C_{similar}$ … 類似度による評価
 W_c … 接続度による評価の重み
 W_s … 類似度による評価の重み

6 解析の流れ

例. 未開 (連体詞性名詞) の 分野 (関係名詞) の 開拓 (サ変名詞)

を用いて上の式の評価値の計算式に基づいて実際に解析を行なってみる。考えられる句構造は、2通りで図1、2のようになる。類似度の評価としては、まず、字面+字面の一致で「未開の分野」が、品詞+字面の一致で「(品詞)の開拓」も検出され、それぞれに評価の点1.0,0.8が与えられる。接続強度の評価は、表3より「未開の分野」「未開の開拓」「分野の開拓」に $5+3=8, 5+2=7, 3+2=5$ のそれぞれ評価を与える。各評価に重みを加えそれぞれの構造のトータルの評価値を決定する。今回は、仮の重みとして $W_c=W_s=1.0$ を与えた。結果、評価値の高い評価1の構造が正しいと推測できる。

表3. 接続強度

	左側接続強度	右側接続強度
連体詞性名詞	1	5
関係名詞	3	3
サ変名詞	2	1

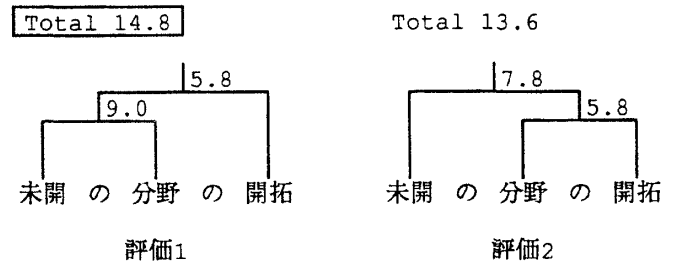


図1: 各構造への評価

7 今後の課題

現在のところ、重みのつけ方は検討中である。この重みのつけ方により解析の正解率はかなり変わってくるものと思われる。今回は、データベースの情報として、名詞の意味的な情報を取り入れていないが、これを加えることによりさらに解析精度の向上が期待できる。

8 まとめ

本稿では、日本語名詞句の「の」で結ばれた名詞句に限定し、その句構造解析を行なうにあたり、接続強度、「名詞+の+名詞」型名詞句の用例を利用する方法を提案した。今後は、各評価の点数のつけ方を検討し、新たな評価の導入も考える必要がある。

謝辞

「EDR 日本語共起辞書」の使用を許可された日本電子化辞書研究所に深謝いたします。

参考文献

- [1] 高橋、佐野、宍倉、前川、宮崎: 頑健性を目指した日本語形態素解析システムの試作、自然言語処理における実働シンポジウム論文集、PP.1-8 (1993)
- [2] 宍倉、宮崎: 構成要素の統語・意味的制約を利用した日本語名詞句解析、信学技報、NLC94-49, PP.41-48 (1995)