

連用・連体接続に基づいた 係り受け構文解析の評価*

1 R - 7

細井 貴晴 小島 丈幸 山口 昌也 乾 伸雄 小谷 善行
(東京農工大学 工学部 電子情報工学科)

1 はじめに

われわれは、日本語文の文節の体言性・用言性に基づく文法系を提示し、体系的な文法規則を作成した [3]。

日本語文の文節は、二つの属性をもつと考えた。一つ目は、文節そのものの性質で、ある文節は体言的か用言的な働きをになうと考える。二つ目は、接続の仕方、ある文節からある文節への係り方には、連用接続か、連体接続、もしくは終止接続があると考えた。

この考えに従った制約規則を定義し、係り受け構文解析をおこなう。そして、得られた結果の定量的な評価をおこなう。また、解の候補を絞るために、解析結果のうち曖昧な解を検証する。

2 制約規則の記述

文節の種類、制約規則を次のように定義した。また、図1に文節の種類における係り受け関係を示す。

- $A, B, C \in \{ \text{体言}, \text{用言}, \text{終止} \}$
- $s(A, B) = \text{文節}$
- $t(A, B) = \text{形態素の列}$

$s(A, B) \rightarrow s(C, A), s(A, B), \{ A \neq \text{終止} \}$.

$s(A, B) \rightarrow t(A, B)$.

3 文法規則の獲得

3.1 文節の認識

制約規則に見合う文節の種類を作成を用例に基づいておこなう。品詞の情報や係り受け関係の文法規則を獲得するためのデータは、次の3.1.1節に示すものを用いた。

*Evaluation of syntactic analysis based on link of declinable words and link of uninflected words, Takaharu HOSOI, Takeyuki KOJIMA, Masaya YAMAGUCHI, Nobuo INUI, Yoshiyuki KOTANI
Tokyo University of Agric. and Tech., Dept. of Computer Science

3.1.1 文節の形を決めるデータ

文節の形を決めるデータは、[1][2]のものを用いた。それらは、毎日新聞の記事を京都大学の形態素解析システム(JUMAN)で形態素解析したものを構文解析システム(KNP)で自動解析を行い、その結果を人で修正されたものである。よって、本研究での品詞の区別は、JUMANの品詞体系に基づいている。

3.2 規則生成のアルゴリズム

文法規則を生成させるアルゴリズムをしめす。

- 係る側
 - 名詞か接尾辞に続く文節 → 体言○○句
 - 動詞か形容詞に続く文節 → 用言○○句
 - それ以外 → 終止○○句
- 受ける側
 - 体言的である → ○○連体句
 - 用言的である → ○○連用句
 - 終止的である → ○○終止句

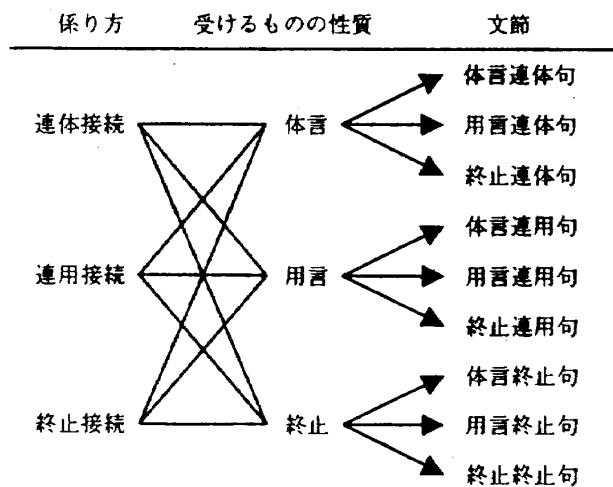


図1: 文節を成す係り受け関係図

3.3 生成された規則

生成された規則の例とその規則の数を表 1 に示す。

表 1: 獲得された規則

文節の種類	構成要素 (例)	数
体言終止句	名詞 + 助詞 + 句点	58
体言連用句	名詞 + 助詞 + 読点	97
体言連体句	名詞 + 助詞 + 読点	83
用言終止句	動詞 + 句点	50
用言連用句	動詞 + 助詞 + 読点	84
用言連体句	動詞 + 助詞 + 読点	50

4 構文解析

4.1 解析の流れ

構文解析をおこなうシステムの流れをしめす。

1. ある文を入力する
2. 文節を認識する
3. 制約規則に当てはめる
4. 解析結果を出力する

4.2 解析例

単純な文における解析結果をしめす。

1. “草食恐竜の王者は、アルゼンチノサウルスだ。”
2. 草食恐竜の → 体言連用句
王者は、 → 体言連体句 体言連用句
アルゼンチノサウルスだ。 → 体言終止句
3. 体言連用句 → 体言連用句 + 体言連用句
体言終止句 → 体言連用句 + 体言終止句
4. 体言終止句 (体言連用句 ((体言連用句 (草食恐竜の)(体言連用句 (王者は、))) 体言終止句 (アルゼンチノサウルスだ。)))

4.3 評価検討

4.3.1 解析文

次の二つの文で、解析をおこなった。

1. 自民党単独政権に終止符を打った事実は、歴史に残る。
2. 前線が活発で、九州は雷を伴った強い雨が降る。

4.3.2 解析結果

曖昧な解析結果や明らかに違う解析結果をも解の候補として出力するので、その解の候補を絞るために、また、制約を加えるために解析結果の考察をおこなう。まずは、得られた解の数を次の表 2 にしめす。

表 2: 解の数

解析文 I	解析文 II
544	568

4.3.3 解の候補を絞る

解の候補を絞るために得られた結果をもとに制約を考える。

- (a) 助詞の情報を細かくする
- (b) 複合語を複合語として認識させる
- (c) 句読点の情報をいれる

4.3.4 新たなる規則の生成

解の候補を絞るために、助詞の情報を細かくして規則を獲得した。助詞は、格助詞・副助詞・接続助詞・終助詞にわけられる。

次に複合語を複合語として認識させるために「名詞+名詞」の結合を強める。

そして、句点・読点で文節が切れるようにする。

以上にしめた制約を導入して、新たな規則を生成して解析をおこなう。

5 おわりに

体系的な構文規則を提案した。そして、二つの文において、解析をおこなった。

また、この構文解析における評価を定量的におこない、今後の指針となる曖昧な解の検証をおこなった。

参考文献

- [1] 京都大学テキストコーパス Version 1.0, 1997.
- [2] コーパス作成の作業基準 Version 1.4, 黒橋禎夫, 斎藤由衣子, 坂口昌子, 1997.
- [3] 細井貴晴 ほか: 連用連体接続に着目した係り受け構文解析, 情報処理学会第 56 回全国大会講演論文集 (2) pp.195-196.