

2ZA-01 文節係り受け解析済コーパスを利用した文生成

小山正太 小島丈幸 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

文節係り受け情報を利用して、ユーザ入力文を脚色し出力するシステムを作成した。本システムでは文節係り受け解析済コーパスから獲得した係り受け知識を利用して、入力文に適当な文節を付け加えることによって文を生成する。コーパスから付加する文節の情報を得るので、コーパスを変えることによって生成文の趣向を変えることができ、コーパスの種類に依存しない。また、文節係り受け情報だけを利用しているので、簡単な構造で文生成が行え、同じ入力文から様々な出力文を生成することができる。しかし、その生成文は不自然になるものがある。簡単な文生成の構造のまま、生成する文の質を高めるために表層格情報や同一語に係る文節同士の共起情報を利用する手法を考えた。表層格情報や共起情報も、文節係り受け情報同様コーパスから得る。本論文ではその手法について述べる。そして、本方法によって得られる文の質を評価する。

2. 文の生成

文の構造は用言を中心にして考える。文を生成するに当たって、文節を付加する方法を、用言に係る場合と用言以外に係る場合との二つに分けて考える。

*A Sentence Generation System using Corpus With Bunsetsu Dependency,

Shota KOYAMA, Takeyuki KOJIMA, Nobuo INUI, Yoshiyuki KOTANI,

Tokyo University of Agri. and Tech., Dept. of Computer Science.

2.1. 文節付加の方法

用言に付加する文節を選択する場合は、係る文節同士の共起情報と表層格情報の二つを用いる。用言以外に付加する文節の選択には文節同士の共起情報を用いる。

付加された文節にさらに係る文節が、文節係り受け情報があれば、また同じように文節付加を行う。この繰返しによって入力文から文生成を行う。

2.1.1. 表層格情報の利用

付加する文節の選択に表層格情報を用いた。一つの用言に同一の表層格に係ると文が不自然になる。このため同じ用言に同一の表層格を付加しないようにする。

2.1.2. 同一語に係る文節同士の共起情報の利用

もう一つの方法として、共起情報を利用した。同一語に係る文節同士は、お互いに意味関係があり、その共起関係を文生成に利用することで、より自然な文が得られる。例 1 のように「買う」に係るという点で、「野菜を」と「鉛筆を」の両文節とも付加する文節としては適切である。しかし「買う」に係る文節同士の意味関係をみると、「八百屋で」と「野菜を」の並びは自然だが、「八百屋で」と「鉛筆を」の並びは不自然になる。そこで、このような一つの単語に係る文節同士の共起情報を用いることによって、単語に付加する際の適当な文節を選ぶ手段とする。

例 1: 「買う」に係る文節

○: 八百屋で 野菜を 買う

×: 八百屋で 鉛筆を 買う

2.1.2.1. 共起情報の利用

付加文節の選択に利用する共起情報として欲し

い情報は、同一語に係る文節の並びである。この文節の並びとは、例 1 の「八百屋で野菜を」のような同一語に係る文節をつなげたものである。この同一語に係る文節の並びをコーパスから獲得した。

2.1.2.2. 共起情報の情報形式

共起情報の利用の方法では、入力文に付加する文節を選び、その文節を含む同一語に係る文節の並びを選ぶ。その文節の並びを、入力文に付加する文節の情報にする。このため、共起情報の情報形式は、ある文節からその文節が含まれる文節の並びを復元できるものが好ましい。そこで例 2 のような形式にした。

例 2：共起情報の情報形式

(文節, 前にくる文節, 後にくる文節)

* : 終端記号

「買う」に係る文節の並びが「私は八百屋で野菜を」の場合の情報はこのようになる。

(私は, *, 八百屋で)

(八百屋で, 私は, 野菜を)

(野菜を, 八百屋で, *)

2.3.1.2. 共起情報の処理例

「買う」に係る文節の「野菜を」が選択された場合、例 2 の情報から、前にくる文節の情報を利用すれば「八百屋で」という文節が選択できる。同じようにして「八百屋で」から「私は」という文節が選ぶこともできる。その結果「私は八百屋で野菜を」という「買う」に係る文節の並びを復元することができる。同じように、「八百屋で」や「私は」という文節からでも同様の結果を導くことができる。

3. 実験

付加する文節の選択になにも利用しない場合、表層格情報を用いて文生成を行う場合、さらに係る文節同士の共起情報を用いて文生成を行う場合

で比較をする。

実験は次の条件のもとで行った。

- ・コーパス内の用言 100 語を使用する。
- ・文節係り受け情報は同じコーパスから獲得してものを使う。
- ・出力文をみて、文法に合っているかどうかで判断する。

3.1. 実験結果

実験結果は次のようになった。

	正解	不正解
なにも利用しない	37%	63%
表層格情報だけ利用	51%	49%
共起情報も利用	68%	32%

3.2. 考察

表層格情報を利用すると正解率が 26%上がった。共起情報を利用すると、なにも利用しない場合より 31%、表層格情報を利用した場合より 17% 生成文の質が向上することが示された。また、「良い天気の日、どしゃぶりの雨の中を歩いた」のように文法的にあっているが、不自然な文が生成されることがあった。これは、同一語に係る文節同士の共起情報は成り立っているが、文節にさらに係る文節との関係をみていないせいである

4. おわりに

本稿では、文節係り受け解析済みコーパスを利用した文生成システムでの、表層格情報と共起情報の利用による生成文の質の向上を目指した。

実験の結果から、表層格情報と共起情報の利用で文生成に効果があることがわかった。今後の課題として、さらに係る文節の情報も見るようにする。

参考文献

- [1] 長尾 真：自然言語処理，岩波書店，1996
- [2] 亀田：軽量・高速な日本語解析ツール『簡易日本語解析系 Q_J P』，言語処理学会第 1 回年次総会，1995