

SL-TRANSにおける文節候補の削減

4G-6

- 係り受け関係を用いた文節候補選択 -

柿ヶ原 康二 森元 逞

ATR自動翻訳電話研究所

1. はじめに

本稿では、音声言語日英翻訳実験システム SL-TRANS の文節候補削減処理について説明する。文節候補削減部は、音声認識部の認識結果である文節ラティスに対して、その候補数を削減し、言語解析の負荷を軽減する。文節候補の削減の方法として、文節間の係り受け情報を用いている。

音声認識結果が文節ラティスになっている場合に係り受け関係を用いて最適な文節候補を選択する手法がいくつか提案されている[1]-[4]。これらは最適文節候補を効率良く求めるものであり、係り受けの曖昧性が高い場合に有効である。しかしタスクが限定されている場合、係り受けの曖昧性は低く、無駄な探索が多くなる。また、文節ラティスをそのまま言語解析部の入力とする文節の組み合わせ数が膨大なものとなり、満足のいく処理時間が得られない。すなわち、比較的簡単な手法で文節候補を削減し、それを言語解析部の入力とすることが現実的である。以下で、係り受け情報による文節候補削減処理について述べる。

2. 係り受けによる文節候補削減処理

2.1 係り受けのレベル

文節ラティス内の文節に対して係り受け関係を調べる際に、表1に示す係り受け関係のレベルを用いる。それぞれのレベルの意味は以下の様である。例を表2に示す。

係り受け関係をサーチする場合、係り受け関係のレベルの順位の高いものから優先的に係り受け関係を付ける。例えばある文節を係り元の文節として係り先の文節をサーチする場合、係り受けのレベルの順位1からサーチを開始する。係り受け関係が付けられない場合は順位2でサーチする。係り受け関係が成立するまで順位を下げ、成立すれば係り先の文節のサーチは終了する。

同じ係り受け関係のレベルで多数係り受け関係が成立する時は、後で説明する係り受けの整合度の最も高いものに係り受け関係を付ける。

表層表現; 対話文データの表現をそのまま用いる。
正規表現; タスク内で同等の内容を示すものを共通の単語で表現したもの。活用語の場合はその終止形。

素性表現; 個々の単語を意味素性で表現したもの。

表1 係り受け関係のレベル

レベル	係り元文節	係り先文節	順位
表層表現	表層表現	正規表現	1
	表層表現	だ文表現	2
正規表現1	正規表現+付属語	正規表現	3
	正規表現+付属語	だ文表現	4
素性表現1	素性表現+付属語	素性表現	5
	素性表現+付属語	だ文表現	6
正規表現2	正規表現	正規表現	7
	正規表現	だ文表現	8
素性表現2	表層表現	素性表現	9
	表層表現	だ文表現	10

表2 係り受け関係のレベルの例

レベル	表現	表現
表層表現	受付に	登録用紙を
正規表現+付属語	事務局+に	用紙+を
素性表現+付属語	human+に	con/doc+を
正規表現	事務局	用紙
素性表現	human	con/doc

2.2 係り受けの整合度[5]

同じ係り受け関係のレベルで多数係り受け関係が成立する時は係り受けの整合度 $P(x,y)$ を調べる。それには係り受けの出現頻度 $F(x,y)$ 、係り受けの距離 $D(x,y)$ 、各文節の認識スコア $S(y)$ を考慮する。現在は次式により整合度を求めている。

$$P(x,y) = F(x,y) - W1 * D(x,y) + W2 * S(y) \quad (Wは定数)$$

2.3 係り受け辞書

係り受けによる文節候補削減を行うための辞書データとしては、『国際会議の参加に関する問合せ』のタスクで収集している対話文から抽出した係り受け関係を用いている。現在、約40000語の対話文からのべ9000個の係り受けデータを作成している。

2.4 係り受け関係のサーチ[5]

文節間の係り受け関係をサーチする場合に、その方法として次の2種類のサーチが考えられる。

(1) 右方向サーチ: 文節 i に右接する部分文節ラティスに対して、文節 i と係り受け関係が成立する可能性の最も高い文節 j をサーチする。

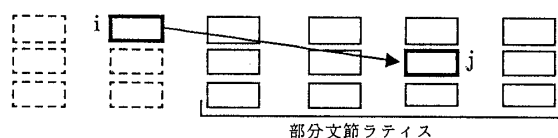


図1 右方向サーチ

(2)左方向サーチ:図2に示されるような係り受け関係のある文節に挟まれている部分文節ラティスに対して、部分文節ラティスの右端の文節候補は、部分文節ラティスに右接する文節 j_3 が部分文節ラティスに左接する文節 i の係り先の文節 j_1 に到達するまでの係り受け関係のある文節 $j_h(1 \leq h \leq 3)$ のいずれかに係る。

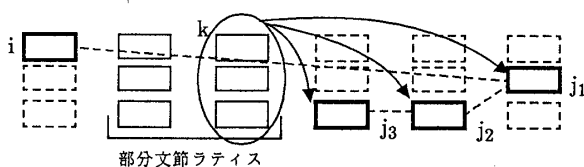


図2 左方向サーチ

2.5 文節候補選択処理手順

上で説明した係り受けサーチの方法を用いて文節候補を選択する手順を以下に説明する。

- (step1) 文節ラティスからサーチのスタートの文節を選択する。これは認識スコアの最も高いものから順に選択する。
- (step2) スタートの文節を係り元の文節として右方向サーチにより文末の文節まで係り受け関係の成立する文節系列をサーチする。
- (step3) 係り受け関係の付けられていない部分文節ラティスに対して、文末に最も近いものから左方向サーチにより係り受け関係の成立する文節をサーチし、係り受け関係の付けられていない部分文節ラティスがなくなるまで繰り返す。

スタートの文節を次に認識スコアの高いものに変えて以上の手順を繰り返す。

3. 文節候補削減結果

サンプル会話に対して文節候補の削減処理を行った結果を以下に示す。

図3は係り受けのレベルごとの文節候補削減の可能な文の数の割合と、その時の正解の文節の順位を示している。係り受け関係のレベルを下げると係り受け関係の成立する組合せが増え、文節候補削減の可能な文の数が増えているが、正解の文節の順位は下がっている。

図4は係り受け関係のレベルを制限した場合の平均文節数の変化を示している。Aは第1位の文節認識率が90%、第2位までが97%、Bは第1位の認識率が2%、第2位までが9%、第3位までが24%の認識結果に対して文節候補を削減した結果である。ABともに第5位までの文節認識率は100%である。Aの場合は認識率が高いので文節候補削減の効果はほとんど見られないが、Bのように認識率が低い場合では文節数の減少が顕著である。係り受け関係のレベルを分けて優先順位を設けた効果が現れており、本手法の有効性が示されている。

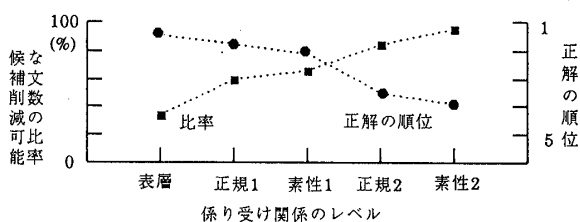


図3 係り受けレベルと絞り込みの関係

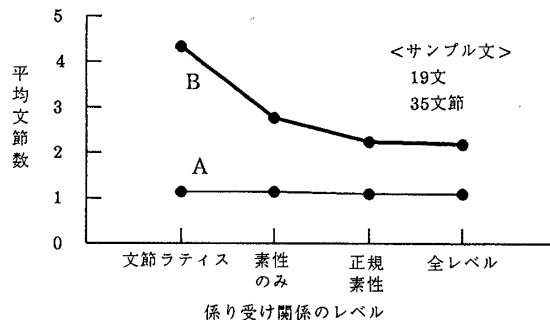


図4 係り受けレベルと平均文節数の関係

4. 問合せ処理

上で説明した文節候補選択処理において文節間の係り受け関係のサーチを行った結果、係り受け関係の付けられない文節候補群が残る場合がある。このときその文節候補群の中に正解の文節が存在している可能性は低い。SL-TRANSではこのような場合にはユーザーに問合せを行い正しい文節の入力を行う。問合せ処理の必要な文節ラティスの例を図5に示す。

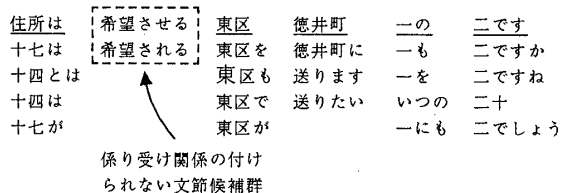


図5 問合せの例

5. おわりに

音声言語日英翻訳実験システム SL-TRANS の文節候補削減処理について述べた。係り受け情報を用いることにより、比較的簡単な手法で文節ラティスの候補数を削減することが可能である。

謝辞

本研究の機会を与えて下さるとともに適切な助言を述べられた樽松明社長に感謝する。また御討論いただいた自動翻訳電話研究所の諸氏に感謝する。

参考文献

- [1] 尾関:「多文節間の係り受け整合度に基づき最適文節列を選択する多段決定アルゴリズム」信学論J71-D(1988-04)
- [2] 中川他:「音節標準パターンと逆時間向き係り受け解析法を用いた日本語文音声の認識」信学論J70-D(1987-12)
- [3] 好田:「文節ラティス上で最適文節列を選択するためのアルゴリズム」信学技法SP86-72(1986-12)
- [4] 松永他:「係り受け規則により候補を限定した文節音声の認識」信学技法SP87-29(1987-06)
- [5] 柿ヶ原他:「係り受け関係を用いた文節候補選択についての一考察」情報処理学会第38回全国大会(1989-03)