最終原稿
012014　文型パターン記述における表記表現の揺らぎ吸収と様相時制の汎化

計算機工学講座 池原研究室　　金澤 佑哉

1　はじめに

日英機械翻訳を実現するために文型パターン辞書が構築されている [1]．文型パターンは，字面，変数，関数，記号で記述されており，パターンマッチングにより入力文の解析を行うものである．文型パターンの汎用性を高めるために，記述要素の効果的な汎化が課題となる．

そこで，本研究では，日本語パターンにおける「時制・相・様相を表す関数の汎化」，および，「要素の位置変更可能の指定」についての効果を，「文型パターン拡大率 η」[2] を用いて，定量的に評価を行い，効果的な汎化であるか確認する．

2　パターン辞書の汎化

2.1　基準パターン辞書

本研究は，[1] のうち文法・単語レベルの文型パターン辞書（122,619 パターン）を基準パターン辞書とする．このパターン辞書を対象に汎化を行う．

2.2　時制関数の汎化

時制関数とはパターンに時制の情報を与える関数であり，過去関数「$.kako$」と未来関数「$.darou$」がある．時制関数が記述されていないパターンは〈現在〉の時制であることが多い．本研究では全ての時制に適合する「自由時制関数」を定義した．

自由時制関数をパターンに付加する方法は 2 つある．1 つは，パターンに記述されている時制関数を自由時制関数に置換する方法である．もう 1 つは，述語付属語部分のうち，時制表現の許される所 [3] に自由時制関数を挿入する方法である．こうして，時制関数を汎化したパターン辞書（自由時制パターン辞書）を作成する．

(汎化前)$N1$ を $V2.kako$ と $V3$。
(汎化後)$N1$ を $V2\sharp1[.kakodarou]$ と $V3\sharp2[.kakodarou]$。

2.3　相・様相関数の汎化

相・様相関数とはパターンに相および様相の情報を与える関数であり，37 種類が定義されている．基準パターン辞書に記述されている相・様相関数の出現頻度を調べ，出現頻度が 1,000 以上の関数 (10 種類) について，関数に任意記号を付けて汎化する．こうして，それぞれ汎化した関数ごとに新パターン辞書を作成する．

(汎化前)$N1$ は $V2.teiru.hitei$。
(汎化後)$N1$ は $V2\sharp1[.teiru]\sharp2[.hitei]$。

2.4　位置変更可能記号の無効化

位置変更可能記号とは出現する位置が変わっても文型パターンとしての意味が変わらない表現要素について，出現できる位置を指定する記号である．

基準パターン辞書には既に位置変更可能記号がパターンに挿入されている．本研究では，基準パターン辞書から位置変更可能記号を取り除いたパターン辞書を作成し，性能の降下を調べる．

(無効化前)$\$1\hat{}\{N1$ は $\}N2$ へ $\$1V3$。
(無効化後)$N1$ は $N2$ へ $V3$。

3　汎化および縮退の効果の測定

3.1　評価パラメータおよび測定方法

本研究の評価パラメータは [2] を参照して，「文型再現率 $R1$」からみた「文型パターン拡大率 η_{R1}」と「平均適合パターン数 N」からみた「文型パターン拡大率 η_N」を用いる．$R1$ は「入力文に対して適合文型パターンが存在するかどうかを文単位で辞書を集計したもの」を表す．N は「入力文に対する適合文型パターン数の平均値」を表す．η は「基準パターン辞書に対して評価対象のパターン辞書が，何倍のパターン数に相当するか」を表す．

本研究ではさらに，η_d を導入する．η_d は「パターンの記述量の違いに基づく拡大率」を表す．η_{R1} と η_N は実測した値だが，η_d は予め予想がつく値である．

2 章で作成したパターン辞書の各 η を調査し，汎化の効果が高いもの調べる．

3.2　測定結果

結果（表 1）より，汎化の効果が最も高いものは時制・相・様相を同時に汎化したパターン辞書だとわかった．

表 1: 測定結果

パターン辞書	η_d	η_{R1}	η_N
基準パターン	1.00	1.00	1.00
(1) 位置変更可能記号を無効化	0.69	0.98	0.95
(2) 時制関数を汎化	2.65	1.53	1.36
(3)[$.teiru$]	1.08	1.19	1.10
(4)[$.rerurareru$]	1.06	1.06	1.08
(5)[$.dadantei$]	1.06	1.02	1.02
(6)[$.hitei$]	1.06	1.05	1.02
(7)[$.teinei$]	1.04	1.03	1.03
(8)[$.meireigotekudasai$]	1.02	1.05	1.02
(9)[$.you$]	1.02	1.02	1.02
(10)[$.suiteirashii$]	1.01	1.07	1.01
(11)[$.sase$]	1.01	1.01	1.05
(12)[$.tekuru$]	1.01	1.01	1.02
(13) 出現頻度が 1,000 以上の相・様相関数を同時に汎化	1.48	1.49	1.40
(14) 時制関数と出現頻度が 1,000 以上の相・様相関数を同時に汎化	4.75	2.15	2.29

4　考察

辞書 (3) から辞書 (13) は予想した拡大率 η_d と実測した拡大率 η_{R1} と η_N との値が近い．しかし，辞書 (1)，辞書 (2)，辞書 (14) は値がかなり異なっていた．

辞書 (3) から辞書 (12) を同時に汎化したものが辞書 (13) であるが，η_{R1}，η_N はそれぞれ辞書 (3) から辞書 (12) の η の積をとると辞書 (13) の η に近いことが判明した．

辞書 (2) と辞書 (3) から辞書 (12) を同時に汎化したものが辞書 (14) であるが，辞書 (14) の η_{R1}，η_N より，時制・相・様相の汎化はかなり効果的な汎化だと考えられる．

5　おわりに

本研究では「要素の位置変更可能の指定」,「時制・相・様相を表す関数の汎化」をする場合としない場合についての被覆率調査を行い，η を用いて定量的に評価した．測定結果より，「要素の位置変更可能の指定」,「時制・相・様相を表す関数の汎化」をすることは，パターンの汎用性を向上するために有効だとわかった．

今後は，単語レベルだけではなく，句レベル，節レベルの文型パターンの汎化を行う予定である．

参考文献

[1] 池原, 阿部, 徳久, 村上:非線形な表現構造に着目した重文と複文の日英文型パターン化, 自然言語処理,11(3),pp.69-95,2004.
[2] 遠藤, 徳久, 村上, 池原:文型パターンにおける任意要素の記述方法とその効果, 言語処理学会第 11 回年次大会,2005(発表予定).
[3] 南不二男:現代日本語文法の輪郭, 大修館書店,1993.