

帰納的学習を用いた自然な日本語文生成手法の評価

尾崎正行 荒木健治 栃内香次

北海道大学大学院工学研究科

〒060-8628 札幌市北区北13条西8丁目

E-mail:{masayuki,araki,tochinai}@media.eng.hokudai.ac.jp

機械翻訳や対話処理などにおいて計算機が生成する日本語文には不自然なものが数多く存在する。このことはユーザにとって不満が残るところである。そのような状況をふまえ、我々は、帰納的学習を用いることにより計算機によって生成される不自然な日本語文をより自然な日本語文に変換する手法を提案している。我々は提案手法の有効性を確認するために実際に実験システムを構築し、評価実験を行った。実験を行った結果、自然な文の割合は32.47%から48.03%へと向上し、本手法の有効性を確認することができた。本稿ではまず最初に、我々の提案手法の概要を説明し、その後実験結果を通して本手法の考察をしていく。

Evaluation of Natural Japanese Sentence Generation Method Using Inductive Learning

Masayuki Ozaki Kenji Araki Koji Tochinai

Graduate School of Engineering, Hokkaido University

N13-W8, Kita-ku, Sapporo 060-8628, Japan

E-mail:{masayuki,araki,tochinai}@media.eng.hokudai.ac.jp

A lot of unnatural sentences exist in the Japanese sentences which the computer generates in the machine translation and the dialogue processing, etc. The user has dissatisfaction for it. From the point of view of this situation, we previously proposed the method which convert an unnatural Japanese sentence generated by computer into its natural Japanese sentence with inductive learning. In order to confirm the effectiveness of our proposal method, we actually constructed the experiment system and carried out the evaluation experiment. The ratio of a natural sentence improved from 32.47% to 48.03%. Therefore we could confirm the effectiveness of our proposal method. In this paper, first of all, we explain the outline of our proposal method, and consider them through the experiment results.

1 はじめに

近年、機械翻訳システムや対話処理システムなどの普及に伴い、計算機上で日本語文を扱う場面が多くなってきている。しかし、計算機が生成する日本語文は必ずしも自然なものではなく、不自然な日本語文が生成されることが非常に多い。特に、機械翻訳のソフトは市販されているものがすでに数多く存在する。しかし、それらのほとんどは翻訳が成功した場合でも直訳をしたものが多い。ここでの翻訳の成否は、原文の意味がとれるかどうかということである。直訳の文の多くは日本語として何か不自然さを感じる。これはユーザにとって不満が残るところであり、より自然な訳文が出力される方が好ましいことは言うまでもない。

このような状況をふまえて、我々は、帰納的学習を用いることにより、不自然な日本語文をより自然な日本語文に変換するシステムの構築を目指している。言い換えの研究はこれまでもいくつか行われているが、それらは、ある目的のために文や単語を別の言葉に言い換えるもの[1,2,3]や、訳語を選択するもの[4]であり、正しい文・単語同士の言い換えを行うものが多かった。さらに、自然性を扱う研究においても、他言語の自然性を扱ったもの[5]は存在したが、日本語文の自然性を扱ったものはほとんど存在しない。本研究は不自然な日本語文を自然な日本語文に言い換えるというものであり、従来の研究とは異なっている。その上、従来の手法はあらかじめ人手により文を解析し規則を記述する手法をとっており、対象とする文をある程度絞る必要があったが、本手法は実例から帰納的学習によりルールを獲得するので、実例が用意できればいかなる文にも対応できるものと考えられる。本稿では、初めに本手法の概要を説明し、実験を通して本手法の評価・考察を行う。

2 システムの概要

システムへの入力文は計算機が生成した日本語文を形態素解析したものとする。ここで形態素解析しか行わないのは、入力文の多くは不自然な日

本語文であり、誤りを多分に含むことが予想され、構文解析や意味解析などを行ったとしても正しい解析結果を得られない場合が多いからである。システムはルール辞書を持ち、ルール辞書中には入力ルールと変換ルールが対になり保持されている。ここで入力ルールとは入力文を可能な限り変数で置き換えたもの、正解ルールとは正解文を可能な限り変数で置き換えたものとする。システムは入力文が入ってくると、ルール辞書を検索し、ルール辞書中に入力文とマッチする入力ルールが存在するかどうかを調べる。もし存在した場合は、それに対応する変換ルールに従い入力文を変換する。現在のシステムは文単位のルールしか獲得・

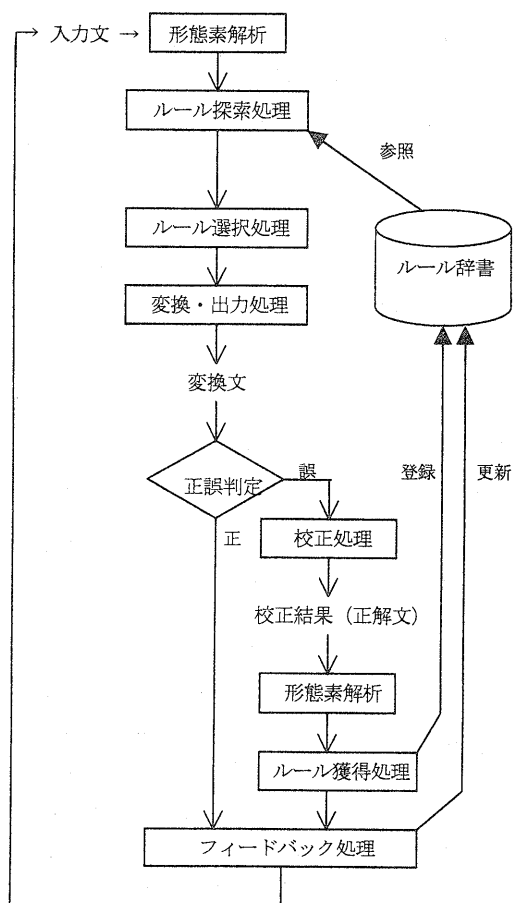


図1: 処理の流れ

保持していないので、一つの文に適用するルールはただ一つである。したがって、もし入力文にマッチする入力ルールが複数存在した場合はルール選択処理を施し、その中から最も確からしい一つのルールを選択する。辞書中に入力文とマッチする入力ルールが存在しなかった場合は正解文（形態素解析済み）を読みこみ、入力文と正解文から新たにルールを作成し、ルール辞書に追加する。処理の流れを図1に示す。

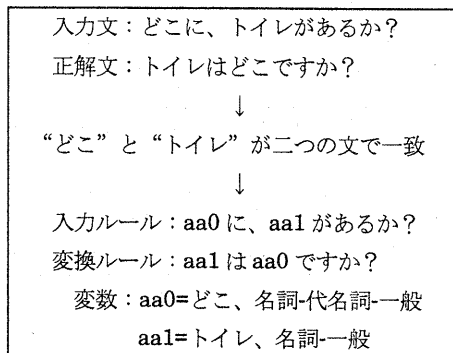


図2：ルール獲得例

3 処理過程

3.1 ルール探索処理部

ルール探索部では、ルール辞書中に入力文にマッチするルールがあるかどうか、ルール辞書中の入力ルールを検索する。入力文・入力ルールともに形態素解析済みであるので、形態素単位での照合となる。入力ルールは変数を含む場合があり、その場合の入力文と変数部分との照合については品詞が一致した場合のみ同一の形態素と考えることとする。文を通して全ての形態素が一致した場合のみ、その入力ルールは入力文にマッチしたとみなす。一つでも不一致の形態素があった場合にはそのルールは不適用となる。その時に注目している入力ルールが入力文にマッチするかしないかが決定すると、次の入力ルールに対して同様の処理を行う。この操作をその段階で保持されている全ての入力ルールに対して行う。

3.2 ルール獲得処理部

入力文と正解文を形態素単位で照合し、同一の形態素があった場合にはその形態素を変数で置き換える。変数に置き換えられた形態素はその品詞・活用形と共に保持される。ただし、変数に置き換えられる対象となるのは自立語のみとする。入力文と正解文を可能な限り変数で置き換えたものをそれぞれ入力ルール・変換ルールとし、この二つは対になってルール辞書中に保持される。ルール獲得例を図2に示す。

3.3 ルール選択処理部

3.1節で説明したルール探索処理部での処理により複数の入力ルールが入力文にマッチした場合に、この処理部により適用するルールを選択する。選択方法は二段階にわかれており、一段階目は変数部分の得点計算、二段階目は入力ルールの尤度計算、となっている。一段階目でルールを選択できなかった場合のみ、二段階目の処理を行う。

3.3.1 変数部分の得点計算

入力文にマッチした入力ルールのうち、変数部分のないものがあつた場合は、その入力ルールをここで適用するルールに決定する。次に、各入力ルールに変数部分が一つ以上存在する場合を考える。各変数部分には今までに適用が成功した形態素が保持されている。この形態素については品詞・活用形・概念情報[6]も同時に保持されている。得点の計算方法は、この変数部分に対応している入力文の形態素が、以前この変数部分に入ったことのある形態素の場合は+2、以前入ったことのある形態素と同一の概念（意味分類）の場合は+1、その他の場合は0とし、この得点分類に従い、各入力ルールの合計点を計算する。この合計点の比較では、変数部分が多いルールの方が有利になってしまうので、合計点を“変数の数×2”でわったものを得点とする（変数の数×2は合計点の満点である）。つまり、入力ルールの得点は、

$$\text{得点} = (\text{合計点}) / (\text{変数の数} \times 2)$$

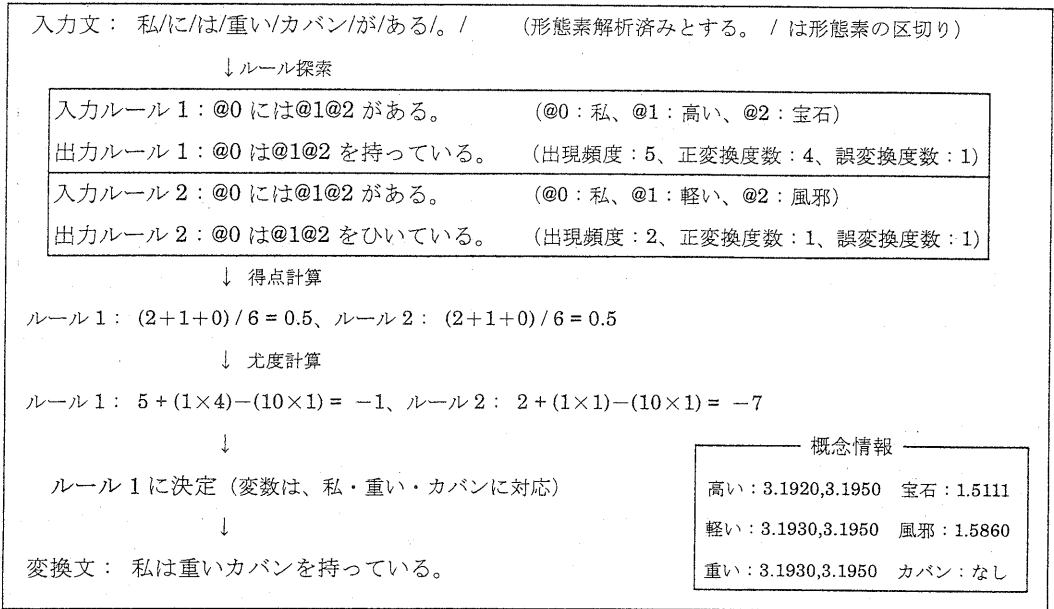


図 3： 全体的な処理の流れの例 ($\alpha=1$ 、 $\beta=10$)

で計算される。

各入力ルールについてこの得点を計算し、この得点の大小を比較する。この段階で最も得点が高いルールが唯一に決まれば、そのルールが今回用いるルールに決定し、ルール選択処理を終了する。もし、唯一に決まらなかった場合は二段階目の入力ルールの尤度計算処理に移る。

3.3.2 入力ルールの尤度計算

ルールの尤度は、以下の式によって計算される。

$$\text{尤度} = \text{AF} + \alpha \text{CF} - \beta \text{EF}$$

ここで、AF はそのルールの出現頻度、CF は正変換度数、EF は誤変換度数、 α ・ β は係数である。

3.3.1 節で述べた変数部分の得点計算の処理で最大の値をとったルールに対して、この尤度計算を行い、尤度が最も高いルールをこの入力文に対して適用するルールとして決定する。

3.4 変換・出力処理部

入力文にマッチした入力ルールが存在する場合は、その入力ルールに対応する変換ルールに従い入力文を変換する。もしマッチする入力ルール

が存在しない場合は、入力文をそのまま出力する。

3.5 フィードバック処理部

ルールが適用され、変換文が出力されたのちに、生成された文に対して人手で正誤評価を行う。自然な場合を正、不自然な場合を誤とする。生成された文が正であった場合、その時に適用したルールの出現頻度と正変換度数を+1する。さらに適用したルールに変数部分があった場合は、その変数部分に対応する入力文の形態素(品詞・活用形も共に)を新たにその変数の値に追加する。生成された文が誤であった場合は適用したルールの出現頻度と誤変換度数を+1し、その時の入力文に対し、3.2 節で述べたルール獲得処理を行う。

最後に、システムの全体的な処理の流れの例を図 3 に示す。この例においては入力文にマッチするルールは、ルール 1・ルール 2 の二つしか存在しないと仮定している。マッチするルールが複数存在するので、得点計算・尤度計算により適用するルールを選択する。その結果、ルール 1 を適用するルールに決定した。

4 評価実験

4.1 実験方法

実験は、旅行者用英会話文 1,703 文を 1 文ずつ翻訳させ、その翻訳結果を形態素解析したものをシステムへの入力文とした。1,703 文の内訳は、ホテルで使用する文 (456 文)、飛行機の機内で使用する文 (345 文)、空港で使用する文 (606 文)、電話で使用する文 (296 文) となっている。また、形態素解析結果を得るために、既存の形態素解析ツール「茶筌」[7]を使用した。システムには各テキストに掲載されている日本語訳を正解文として与え、ルール獲得処理を行う際にはこれを用いた。ルール辞書の初期状態は空とし、3.3.2 節で述べた尤度計算式における係数 α 、 β はそれぞれ、 $\alpha=1$ 、 $\beta=10$ とした。

システムの性能評価は、正変換率・誤変換率・未変換率の三つを用いて行う。それぞれは、以下の式により定義される。

$$\text{正変換率} = \frac{\text{出力が自然であった文数}}{\text{全入力文数}}$$

$$\text{誤変換率} = \frac{\text{出力が不自然であった文数 (ルール適用)}}{\text{全入力文数}}$$

$$\text{未変換率} = \frac{\text{出力が不自然であった文数(ルール未適用)}}{\text{全入力文数}}$$

4.2 実験結果

1,703 文に対して変換処理を施した結果、正変換率は 48.03%、誤変換率は 0.82%、未変換率は 51.15% となった。1,703 文中に占める自然な文の割合は、実験前は 32.47% であったが、実験後は 48.03% に上昇した。100 文単位での各変換率の推移を図 4 に示す。同じく 100 文単位での実験前、実験後の全文に占める自然な文の割合の推移を図 5 に示す。

5 考察

5.1 実験データについて

今回は実験データとして旅行者用英会話文を用いたが、翻訳を必要とする文はもちろん旅行者用英会話文だけではないので、もっと多種多様なデータでの実験を行う必要がある。さらに、本稿の初めにも述べたが、機械翻訳だけでなく対話処理などにおいても計算機が生成する日本語文には不自然なものが数多く存在する。我々の提案手法は機械翻訳のみに対応するものではなく、対話処理などにおいて生成される不自然な日本語文にも対応できるはずであり、それを確認するためにも機械翻訳以外で生成された日本語文についても今後実験を行う必要がある。

5.2 各変換率について

注目すべき点は、誤変換率が低いことである。これは入力文にマッチするルールが見つかりさえすれば、ほぼ間違いなく自然な文を生成することが可能であるということの意味する。つまり、このシステムへ入力をするることによる悪影響がほとんどないわけである。この結果より、我々の提案手法の有効性を示すことができたと考えている。しかし、その一方、未変換率が 50% を越えている。この問題については、現在は文単位のルールしか獲得・適用していないことが原因として挙げられる。今後、句単位などのより短いルールを獲得・適用するようにシステムを改良していけば、未変換率を減少させることが可能であると考えている。

5.3 尤度計算のパラメータについて

本稿では $\alpha=1$ 、 $\beta=10$ としたが、実際には $\alpha=1$ を固定して β を 1~20 まで変化させた 20 通り、 $\beta=10$ を固定して α を 1~20 まで変化させた 20 通り、延べ 40 通りについて実験を行った。しかし、結果は α 、 β によらず全ての場合において同じであった。これは、現在のシステムは誤変換した場合にのみ、入力文にマッチするルールがすでに存在するにもかかわらずルール獲得処理を行う。

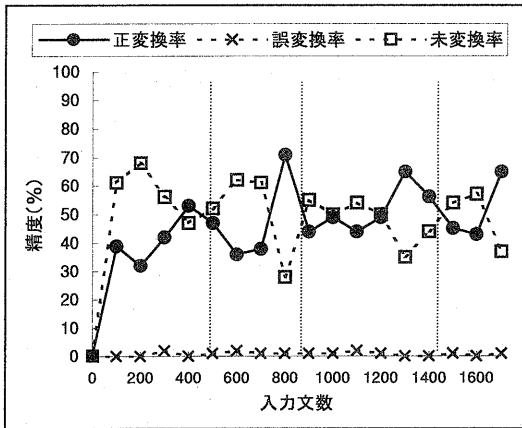


図4：各変換率

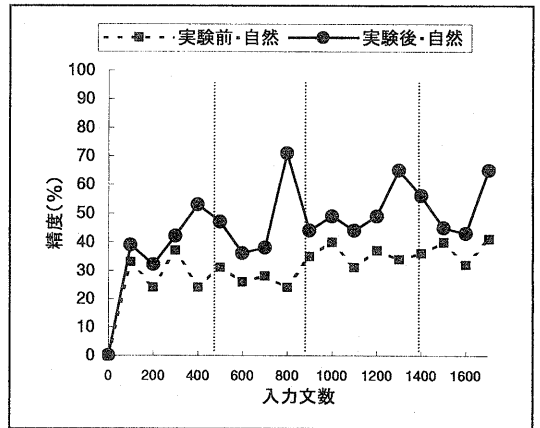


図5：自然な文の割合

その結果、似たルールが複数でき、その後の処理において複数のルールが競合するわけであるが、今回の実験では誤変換率が1%未満であり、似たルールが複数できることがほとんどなかったからである。その上、複数のルールが競合した場合においても、ルール選択処理においては先に変数部分の得点計算が行われるため、その段階でルールが決定することが多かったことが理由として挙げられる。今後、ルールの競合が数多く起こる状態になった場合に、この係数について再度実験・考察をする予定である。

6 おわりに

本稿ではまず最初に、機械翻訳や対話処理などにおいて計算機が生成する不自然な日本語文を自然な日本語文に変換する手法を提案した。その後、提案手法の有効性を確認するために、実際に実験システムを構築し、評価実験を行った。その結果、自然な文の割合は、32.47%から48.03%へと向上した。さらに、誤変換率は1%未満であり、このシステムに入力をするることによる悪影響がほとんどないことを確認することができた。これらの結果は、本手法の有効性を示すものである。

今後は文単位だけでなく、句単位などのより短

いルールを獲得するように改良を加え、より精度をあげるように研究を進めていく予定である。

参考文献

- [1]黒橋禎夫：言語の意味を計算機で扱う，言語処理学会第6回年次大会チュートリアル資料，pp21-28(2000)。
- [2]佐藤理史：論文表題を言い換える，情報処理学会論文誌，Vol40，No7，pp2937-2945(1999)。
- [3]A.Kilgarriff.SENSEVAL：An Exercise in Evaluating Word Sense Disambiguation Programs. In Proc. of LREC, Granada, pp581-588, 1998.
- [4]野見山浩：目的言語の知識を用いた訳語選択とその学習性，情報処理学会，自然言語処理研究会,86-2(1991)。
- [5]山本和英：機械翻訳における自動校正と日中翻訳への適用，言語処理学会第5回年次大会発表論文集，pp21-24(1999)。
- [6]国立国語研究所：“分類語彙表”，秀英出版(1964)
- [7]松本裕治，北村啓，山下達雄，今一修，今村友明。“日本語形態素解析システム『茶筌』version2.0 使用説明書。” Technical Report NAIST-IS-TR99008, 奈良先端科学技術大学院大学, 1999