

212 英日機械翻訳における帰納的学習を用いた自然な訳文の生成

尾崎正行 荒木健治 栃内香次
北大工

1.はじめに

現在、市販されている翻訳ソフトは数多くあるが、それらのほとんどは、原文の意味がとれるという意味において翻訳が成功した場合でも、直訳をしたものが多く、その結果、日本語として何か不自然さを感じる訳文が非常に多いことに気がつく。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、不自然でかたい翻訳結果をより自然な訳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]や、訳語を選択するもの[3]であり、正しい文・単語同士の言い換えであった。それに対し、本研究は不自然な文を自然な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。さらに、従来の研究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、本研究は帰納的学習を用いており、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簡単に説明し、実験を通して本手法の考察を行う。

2.概要

入力文は翻訳結果を形態素解析したものとす。ここで形態素解析しか行わないのは、入力文は翻訳結果であり、誤りを多分に含むことが予想され、構文解析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、辞書中に入力文とマッチする入力ルールが存在した場合は、その変換ルールに従い入力文を変換する。もし、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。

2.1 ルール作成

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・活用形とともに保持される。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。ルールは、入力文・正解文を可能な限り変数で置き換えたもの、変数に置き換えられた形態素・その品詞・活用形ということになる。ルール作成例を図1に示す。

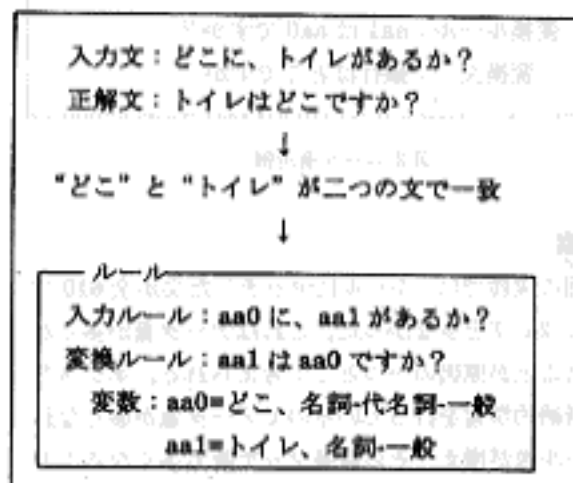


図1 ルール作成例

2.2 ルール照合・適用

入力文にマッチするルールがあるかどうかルール辞書の中の入力ルールを一つずつ見ていく。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。変数部分との照合は品詞が一致した場合のみ同一の形態素と考え、次の形態素に進む。文を通して全ての形態素が一致した場合、その入力ルールは入力文にマッチしたとみなし、変換ルールに従い入力文を変換する。一つでも不一致の形態素があった場合には、その入力ルールはこの入力文には不適用となり、次の入力ルールを見にいき同様の処理を繰り返す。ルール適用例を図2に示す。

3. 実験

旅行者用英会話文[4]の翻訳結果に対して、本システムの評価実験を行った。結果は全610文に対し、ルールにマッチした文が21文、そのうち17文が正しい日本語訳を出力することに成功した。ルール正適用例と誤適用例を付録として最後に示しておく。

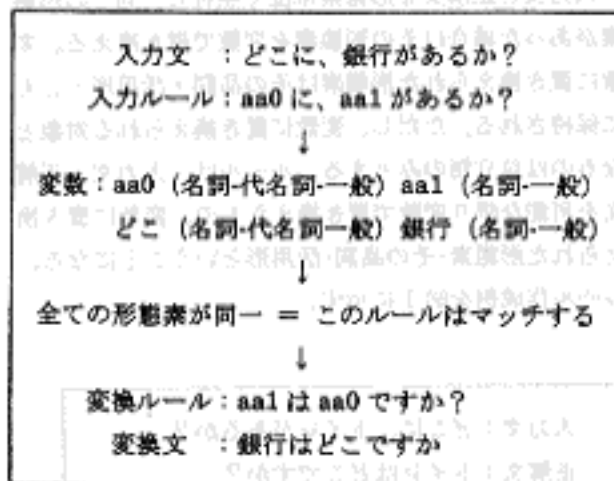


図2 ルール適用例

4. 考察

今回の実験では、ルールにマッチした文が全610文に対し21文と少なかった。これはデータ量が多くなかったことが原因の一つとして考えられる。本システムは帰納的学習を行っているのでデータ量が多くなればルール数が増え、その結果マッチ数も多くなることが当然予想される。したがって、データ量を多くすればこの問題点はある程度解消できるものと思われる。他にマッチ数が少なかった原因として考えられることは、入力文と入力ルールがマッチする条件が厳しかったことである。現在は、字面部分は字面の完全一致が同一の形態素とみなす条件となっているが、この条件では精度は良いがマッチ数が増えないので、この条件をどのように改良していくかが今後の課題である。さらに、今回の実験ではあまり問題にはならなかったが、現システムでは変数部分の照合において品詞の一致を同一の形態素とみなす条件としているが、品詞の一致という条件だけでは不足していると考えられる。その問題を解消するために、将来的にはシソーラスの導入を予定している。

5. おわりに

本稿では、提案手法の概要を簡単に説明し、その後、実験を通して本手法の考察を行った。今回の実験はデータ量が少なかったこともあり、本手法の有効性を明確に示すまでには至らなかったが、数少ないながらも言い換えには成功しており、本手法が有効である可能性は示せたと考えられる。今後はデータ量を増やして再度実験を行い、その結果・考察をふまえて本手法の改良を行う予定である。

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]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).
- [4]地球の歩き方編集室:旅の会話集④ 米語/英語,ダイヤモンド・ビッグ社(1993).

付録

ルール正適用例

入力文: どこに、洗面所があるか?

変換文: 洗面所はどこですか?

入力文: 私は窓を開いてもよいか?

変換文: 窓を開いてもいいですか?

入力文: この席は取られるか?

変換文: この席はあいていますか?

入力文: これは何であるか?

変換文: これは何ですか?

ルール誤適用例

入力文: どこに私は行くべきであるか?

変換文: どこで行くのですか?

入力文: どこに、現在、私はいるか?

変換文: 現在、私はどの辺りにいますか?