

対訳パターンを辞書とする翻訳システム

4K-4

村田稔樹, 杉尾俊之, 坂本仁
 沖電気工業株式会社 関西総合研究所
 mura@kansai.oki.co.jp

1 はじめに

近年、規則ベースの翻訳方式では膨大な規則の保守の難しさが問題となっており [1]、用例ベースの翻訳方式の研究が盛んに行われている。しかし、その多くは、処理すべてを用例ベースで行なっているわけではなく、従来の規則ベースの処理を組み込んで構成されている。

そのため、規則ベースの部分で行なわれた処理が誤った結果を出力した場合、用例を追加しても訳が改善されないといった致命的な問題を抱え持つことになり、用例ベースの本来の利点を生かすことができない。

そこで我々は、基本に立ちかえり、すべての処理を用例ベースで行なう翻訳システムを提案する。

2 用例ベースの翻訳システム

図1のように、現在の用例ベースによる翻訳システムは大きく以下の3つに分類できる。

- (1) 規則ベースの翻訳処理の前処理（またはバイパス）として用例を用いる [2][3]
- (2) 規則ベースの翻訳処理の変換部分のみを用例ベースで行う [4][5]
- (3) 用例ベースを主とするが、それで処理できない場合、規則ベースを呼び出す [6]

このように現在の用例ベースの翻訳システムの多くは、規則ベースのものと併用している。

ところが、これでは各々のモジュールが持つ知識が共有できないだけでなく、整合性が保たれているという保証がない。上記の (1) では、まったく別の翻訳システムが1つの箱に入れてあるだけで、その相互作用はまったくなく、(2) では解析部は規則ベースであるため用例ベースの利点を生かすことができない。また、(3) でも結局は複雑な文は規則ベースで解析するため、規則ベースでの問題点が必ずつきまとう。

Machine Translation System Using Dictionary of Translation Patterns

Toshiki Murata, Toshiyuki Sugio, Masashi Sakamoto
 Kansai Laboratory, Oki Electric Industry Co., Ltd.
 1-2-27, Shiromi, Chuo-ku, Osaka, 540, Japan
 Phone:+81 6 949 5101, Fax:+81 6 949 5108,
 Email: mura@kansai.oki.co.jp

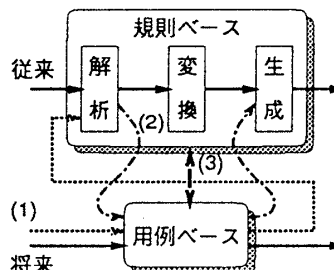


図1: 用例ベースの翻訳システム

ここでは、用例ベースの利点である「用例を追加して学習することによってどのような誤った訳も改善することができる」ということを最初から放棄していることになる。

そこで我々は、あくまで用例ベースのみで翻訳処理を行う方法を提案する。まずは、用例として第3節で述べるような対訳パターンを用いる。この対訳パターンを辞書に持たせ、第4節で述べるように翻訳処理を行なう。このシステムは、従来の規則ベースの処理を一切用いないため、誤った訳が出力された場合、用例を追加することで必ず訳を改善することができる。

3 対訳パターン

原言語のあるフレーズが目的言語のあるフレーズに翻訳できる場合、それらのフレーズの組を対訳パターンと呼ぶ。また、このフレーズの一部分が変わってもその他は同じように訳す場合、その部分は変数(下の例で「~」「…」【{ }】の部分)と呼ぶ。

私はペンを持っている // I have a pen
 ~は…を持っている // ~ have …

また、同じ原言語パターンでも変数にマッチする語句によって訳し分けたい場合がある。その場合、変数部分に具体的な言葉を付加し、古瀬ら [6] のように、翻訳時にそれとの類似度をもとに訳し分けする。

~は…を持っている // ~ have …
 ↓「…」の詳細化
 ~は【カバン】を持っている // ~ have {a bag}
 ~は【家】を持っている // ~ own {a house}

このように、対訳パターンは、知識の表現形式として直接的で分かりやすく、ユーザが作成/保守する上で非常に優れている。さらに、重要なことは、未加工のraw コーパスからでも自動的に取り出すことが可能である [7][8] と考えられる。

4 対訳パターンのみを用いた翻訳処理

4.1 解析

これまで用例ベースだけでシステムを構成できなかったのは、用例ベースでの解析が困難だったためと考えられる¹。確かに、入力文と最も類似した文を検索し、異なる部分をまた同様に検索していくという方法では現実的には不可能に近い。しかし、これは文を検索していたからであって、第(x対訳パターン節で述べた対訳パターンを用いれば実はかなり容易に解析することができる。まず、簡単に解析できるようにするには、従来のルールベースの解析で用いてきた知識(特に辞書)を対訳パターンに変換すればよい。

例えば、

連体詞 → この
名詞 → ペン
名詞句 → 連体詞, 名詞

のような知識は、

この～ // this ～
ペン // pen

という対訳パターンに変換できる²。

このように既存の辞書を対訳パターンに変換して辞書化しておけば、後は、入力文に対して対訳パターンの原言語側を組み合わせることによって解析するだけである。もちろん組み合わせには曖昧性が伴うため、類似度などを用いて古瀬ら[6]のように最尤構造を求めることによって曖昧性を除去する。

例えば、

(a) ～は…を持っている // ～ have …
(b) 私の～ // my ～
(c) ペン // pen

という対訳パターンを組み合わせるとき、(b)の「～」の部分には、(c)の「ペン」をマッチさせ、その結果まとまった「私のペン」を(a)の「…」の部分にマッチさせる。つまり、変数にパターンとパターンを接続させるという意味を持たせる。その結果、接続したパターンは1つのフレーズ(まとまり)を示すことになる。

この変数による接続表現を統語的な面から見てみると以下ようになる。

| 左端 | 右端 | 役割 | 統語的には | 例 |
|----|----|-----------|---------|-------|
| ない | ない | それだけでまとまる | 名詞 | 本 |
| ある | ない | 前と接続する | 動詞など | ～に行く |
| ない | ある | 後ろと接続する | 連体詞・形容詞 | きれいな～ |
| ある | ある | 前後をつなぐ | 接続詞, 助詞 | ～と… |

¹変数を含んだパターンの解析方法としては、丸山の方法[9]があるが、パターンをそれぞれ非終端記号に割り付けてカテゴライズしているため、パターンの柔軟な適用が阻害され、また、類似度などを取り込めないなど、用例ベースの解析としては不十分である。

²この過程で情報が落ちているが、解析をするためにこの情報が不可欠だったかどうかは言い切れない。逆に、これぐらいの情報で解析できるという方が機械学習という面から見ると有用である。

このように、原言語パターンの左端と右端に変数があるかどうか、統語的な役割を分類することになる。

4.2 訳出

対訳パターンの原言語側のパターンを組み合わせると解析し終わったところから、順次目的言語側パターンに置き換えることにより訳出する。

5 おわりに

本稿では、対訳パターンのみを用いて解析・訳出する翻訳システムを提案した。このシステムは、対訳パターンのみを用いて翻訳するため、誤った訳が出力された場合用例を追加することで必ず訳を改善することができるなど、用例ベースにおける利点を十分に生かしたものとなっている。

また、対訳パターンの原言語側を用いて解析し、解析できた部分から直ちに目的語側に変換して訳出することができるので、人間の同時通訳のように1文入力されるのを待つことなく順次訳出することができる。

今後は、このシステムをプラットフォームにして、対訳パターンの記述能力や自動獲得を研究することによって、実用に供する用例ベースの翻訳システムを作成する予定である。

参考文献

- [1] 村田：今、現場ではどんな理論が必要か, 信学会, 日本ソフトウェア科学会「自然言語処理における実動」シンポジウム論文集, pp. 152-154 (1993).
- [2] 浦谷, 加藤, 相澤: AP 電経済ニュースからの定型パターンの抽出, 情処第 42 回全国, pp. 6E-4 (1991).
- [3] Kinoshita, S., Kumano, A. and Hirakawa, H.: Improvement In Customizability Using Translation Templates, in *Proc. of Coling-94*, pp. 25-31 (1994).
- [4] 佐藤: MB T 2 : 実例に基づく翻訳における複数翻訳例の組合せ利用, 人工知能学会誌, Vol. 6, No. 6, pp. 861-871 (1991).
- [5] Watanabe, H.: A Similarity-Driven Transfer System, in *Proc. of Coling-92*, pp. 770-776 (1992).
- [6] 古瀬, 隅田, 飯田: 経験的知識を活用する変換主導型機械翻訳, 情処論, Vol. 35, No. 3, pp. 414-425 (1994).
- [7] Kaji, H., Kida, Y. and Morimoto, Y.: Learning Translation Templates From Bilingual Text, in *Proc. of Coling-92*, pp. 672-678 (1992).
- [8] 北村, 松本: 二言語対訳コーパスからの翻訳知識の自動獲得, 人工知能第 8 回全国 (1994).
- [9] Maruyama, H.: Pattern-Based Translation: Context-Free Transducer and Its Applications to Practical NLP, in *Proceedings of NLPRS'93*, pp. 232-237 (1993).