6Q－2

多段用例翻訳における情緒的表現の翻訳機構

冨樫亮介　宮崎正弘
新潟大学大学院自然科学研究科

1 はじめに

「辞」「詞」「句」「複合名詞」など、複数のレベルの対訳コーパスを用いた、多段用例翻訳方式の研究[1]が行なわれている。これに三浦文法に基づくパーザ[2]を組み合わせ、客体的表現と主体的表現の分離を行なうことにより、人間の情緒的な表現を多く含む文章の翻訳に、高精度な翻訳が可能となることを示す。

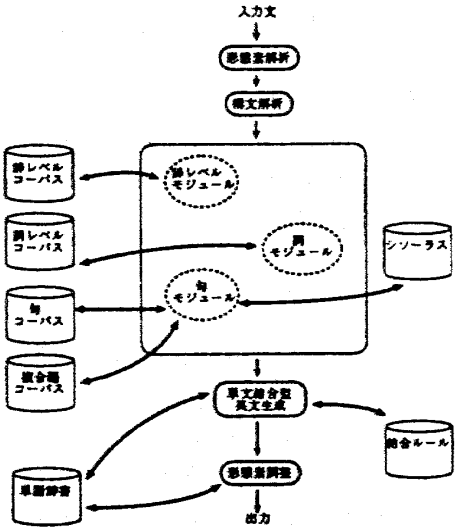

図 1: 多段用例翻訳システムの概要

2 マルチコーパスを用いた多段用例翻訳

図1に示すように、マルチコーパスを用いた多段用例翻訳では、「辞」「詞」「句」「複合名詞」などの文を構成する種々の階層に応じたコーパスを用意し、入力文に対して、複数のコーパスを適応させる。そして、各レベルで得られた対訳を組み合わせて、目的言語に対応する訳文を生成する。この種々のレベルのコーパスを効果的に利用するため、それぞれのレベルに応じた類似評価を行なう、各種のモジュールが存在する。

本稿で述べる情緒的表現の翻訳機構も、マルチコーパスを用いた翻訳システムの一モジュールである。

3 客体的表現と主体的表現の分離

三浦文法によれば一つの文は客体的表現と主体的表現から構成される。主体的表現とは人間の意志や感情を直接表現したものであり、客体的表現とは対象を概念として表現したものである。この考えに基づき、本研究で用いるパーザは主体的表現と客体的表現の分離を行う。加えて、主体的表現のコーパスを準備し、人間の情緒的表現の柔軟な訳出に対応している。

主体的表現を抽出するに当たっては、ほぼ表層的な形態素段階において処理される。

形態素段階では、形態素レベルの段階で主体的表現が分離できるものを分離する。この段階で分離するのは、「すみませんが」「～して下さい」などといった特定の表現と、「～して下さいませんか」「～していただけないでしょうか」などの辞が複合した複合辞表現である。このような複合辞はルールを用いて処理する。辞を分離した後、残った部分はパーザで扱える正規的な表現に書き換えてパーザに渡してやる。また、辞の解析部分のデータは後述の辞レベルコーパスの評価モジュールに渡される。

実際の処理例

図2に入力文「学校へ行く道を教えて下さいませんでしょうか。」が与えられた時の処理を示す。この場合は形態素段階で、辞の部分「て／下さい／ませ／ん／でしょ／う／か／。」を

Translation Mechanism of Emotional Expression in Example-based Translation using Multi-Corpus
Ryosuke Togashi , Masahiro Miyazaki
Niigata University

分離する。「動詞＋て下さい」は依頼を示す特定表現、「ませ／ん／でしょ／う／か／。」はそれぞれ１個の形態素として見れば、肯定・否定・肯定（丁重）・推量・疑問を示すが、複合した場合は「疑問（丁重２）」として抽出する。すなわち形態素段階で分離した辞は「依頼・疑問（丁重２）」となり、この部分は辞レベルコーパスの評価モジュールに渡す。そして、残った部分「学校へ行く道を教え」を「学校へ行く道を教える」と書き換えて、パーザに渡す。

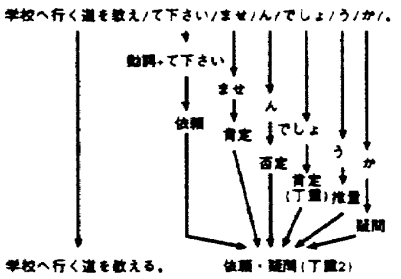

図 2: 形態素からの辞の分離

4 辞を考慮した用例検索モジュール

このモジュールでは構文解析の結果と前述の形態素レベルでの辞の解析データの両方を用いて、コーパスを検索し、対応する英語側の表現素片を抽出し、さらにそれとの差分の素片を求める処理を行なう。

構文解析結果と辞データを組み合わせてコーパスと照合をした場合、ある素片がマッチングしても、その差分もまたマッチングさせ... という処理を繰り返し、いくつかの素片と差分を得る。そして、いくつかの素片の組み合せから、対応する英語の表現素片を得る。

図３の例でいくと、「学校へ行く道を教えて下さいませんか。」は辞の部分「依頼・疑問（丁重）」と「学校へ行く道を教える」の木構造の組み合わせとして扱う。その組合せと類似する日本語側の素片、「～を教える＋依頼・疑問（丁重）」がまず検索され、そこから、差分「学校へ行く道を～」と英語の素片"Colud you tell me ～" を得る。ついで、その差分の類似をとり、そこから、対応する英語側の素片、"the way to ～" と、差分として「学校」を抽出する。

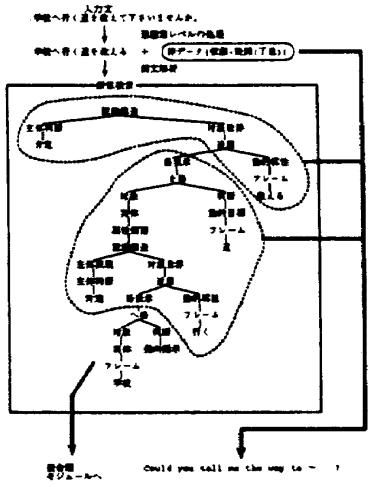

図 3: 辞データと構文解析結果からの用例検索

このモジュールから得られる英語側の素片は "Colud you tell me the way to ～" となる。差分「学校」は複合語レベルのモジュールに渡され、日本語に翻訳される。

5 おわりに

本稿ではマルチコーパスを用いた用例翻訳システムにおいて、三浦文法に基づき、主体的表現と客体的表現の分離をうまく行なうことによって、人間の意志や感情といった情緒的側面を多く含んだ文章を効果的に解析する方式を示した。

現在本方式に基づいた日英間の多段用例翻訳のプロトタイプを試作中であり、今後システムの構築と、コーパスを充実を進めると共に、翻訳品質の総合的評価が必要となる。

参考文献

[1] 池田、宮崎：マルチコーパスを利用した多段階用例翻訳方式、情報処理学会第５３回全国大会、4L-9(1996)

[2] 薮、藤石、宮崎：表現構造と話者の認識対象構造を抽出する日本語文パーザの試作、言語処理学会第３回年次大会、pp.205～208(1997)