

## 論文内容の要旨

博士論文題目 Acquiring Paraphrases from Corpora and Its Application to Machine Translation  
(コーパスからのパラフレーズの獲得とその機械翻訳への適用)

氏名 下畑 光夫

### (論文内容の要旨)

自然言語には、パラフレーズ、すなわち同じ意味を共有する表層的に異なる表現が存在する。同じ意味を様々な表現で表すことができるということは自然言語の表現力の豊かさを示すものであるが、一方で自然言語処理の性能を低下させる一因となる。この対処のために、シソーラスなどの言語資源が構築されているが、それらは汎用的情報を記述しているために、特定分野や特定アプリケーションにとって十分な効果を持っていない。

本論文の目的は、コーパスに基づくパラフレーズの獲得ならびにその機械翻訳への適用である。本論文では、語彙的パラフレーズ獲得と文レベルパラフレーズ獲得の2種類の方式を提案する。両手法は表層的処理に基づいており、コーパス以外の資源を必要としない。本論文では、人手によるパラフレーズの分析、語彙的パラフレーズ獲得、類似文検索、の3点について述べる。

人手によるパラフレーズについては、2種類の研究について報告する。この研究では、(1) 人手によるパラフレーズでは、どのようなタイプのパラフレーズが多く行われるか、(2) 人手によるパラフレーズを機械翻訳に与えた場合にどの程度の効果が得られるか、について検証する。

次に、パラレルコーパスから語彙的パラフレーズを獲得する方法について述べる。提案手法は以下の特長を持つ。(1) 機能語に関連するパラフレーズを獲得できる。(2) 単言語からの観点だけでなく、対訳的観点から見て同義であるパラフレーズが獲得できる。コーパス中に存在するパラフレーズを単一の表現に置き換えることにより、コーパス中のテキストを簡潔化することができる。この簡潔化により、コーパスに基づく機械翻訳の性能を向上させることができる。

最後に、単言語コーパスから類似文を検索する方法(文レベルのパラフレーズ)について述べる。提案手法は、収集が容易である単言語コーパスを用いることと、語彙レベルを超えるパラフレーズを獲得できるという特徴を備えている。3種類の主要な方式について比較実験を行い、類似文検索方式としてN-gramの共通部分に基づく方式を採用している。本手法を以下の2通りの方法で機械翻訳に適用した。一つは入力文の前編集として利用する方法である。与えられた入力文が翻訳不能である場合に、翻訳可能な類似文を検索し、それを機械翻訳に渡すことで翻訳不能文を翻訳することが可能となる。もう一つは、用例翻訳の類似度算出に用いる方法である。本手法は長文や文体の違いにロバストな検索を実現しており、従来手法と比較して広いカバレッジを実現している。

氏名	下畑 光夫
----	-------

(論文審査結果の要旨)

平成16年7月20日に開催した公聴会の結果を参考に平成16年9月14日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

下畑 光夫は、本博士論文において、パラフレーズを利用することによって機械翻訳システムのカバレッジを向上させ、翻訳性能を改善する手法を提案した。また、訳出性能の向上を様々な観点から評価している。まず、本論文では、パラフレーズのタイプを、語彙レベル、構文レベル、および文レベルに分類し、特に機械翻訳の性能向上のために、語彙レベルと文レベルのパラフレーズをコーパスから自動獲得する手法を提案した。

1. 語彙レベルのパラフレーズ獲得では、翻訳結果を共有する対訳コーパスから、前後文脈を考慮した語彙パラフレーズを抽出する方法を示し、その有効性を評価した。
2. 文レベルのパラフレーズ利用では、単語 N-gram に基づいて文間の類似度を定義し、機械翻訳システムにとって翻訳不能な文に対して類似した翻訳容易文を検索し、その結果を翻訳システムに入力することで翻訳システムのカバレッジを拡大することに成功した。また、類似度尺度を用いることで、用例に基づく機械翻訳や統計機械翻訳の性能向上が可能であることを示した。

このようにパラフレーズを用いることによって機械翻訳システムの性能向上を達成した本論文の提案手法は実用的であり、かつ、新規性もあり、機械翻訳技術の分野において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。