日英機械翻訳における利用者登録語の意味属性の自動推定

池原　悟[†]　白井　諭[†]　横尾　昭男[†]
Francis Bond[†]　小見　佳恵[††]

機械翻訳システムを使用して現実の文書を翻訳する場合, 通常, 翻訳対象文書に合った利用者辞書が必要となる. 特に, 高品質翻訳を狙った機械翻訳システムでは, 各単語に対して, 約 2,000 種以上の分解精度を持つ単語意味属性の付与が必要であると言われており, 一般の利用者が, このような精密な情報を付与するのは困難であった. そこで本論文では, 利用者が登録したい日本語名詞 (複合名詞を含む) と英語訳語を与えるだけで, システムがシステム辞書の知識を応用して, 名詞種別を自動的に判定し, それに応じた単語の意味属性を付与する方法を提案する. 本方式を, 新聞記事 102 文とソフトウエア設計書 105 文の翻訳に必要な利用者辞書作成に適用した結果, 自動推定方式では, 専門家の付与した意味属性よりも多くの属性が付与されるが, 40 ~ 80%の再現率が得られることが分かった. また, 人手で作成した利用者辞書を使用する場合と同等の訳文品質が得られることが分かった. 以上の結果, 利用者辞書作成への単語の登録において, 最も熟練度の要求される単語意味属性付与作業を自動化できる見通しとなった.

キーワード:　機械翻訳, 利用者登録語, 意味属性, 自動推定

Automatic Determination of Semantic Attributes for User Defined Words in Japanese to English Machine Translation

Satoru Ikehara[†] , Satoshi Shirai[†] , Akio Yokoo[†] , Francis Bond[†]
and Yoshie Omi[††]

User dictionaries are important for practical machine translation. However it is difficult for users to enter the detailed semantic attributes that a system may require. This paper proposes a method of automatically determining semantic attributes for noun pairs entered by users. The method compares the Japanese and English words with words already in the system dictionary. When this method was applied to words in two user dictionaries (for newspaper articles and software manuals) it generated more attributes than trained humans did, with a recall rate of 40–80%. Evaluation showed that translations using the machine-made dictionary were similar in quality to translations using the human-made dictionary. Thus automatic determination of semantic attributes removes the need for highly trained lexographers to make user dictionaries.

KeyWords: *machine translation, user dictionary, semantic category, automatic determination*

† NTT コミュニケーション科学研究所, NTT Communication Science Laboratories
†† NTT アドバンステクノロジ, NTT Advanced Technology Corporation