

日英機械翻訳用構文意味辞書の 記述精度の向上と作成支援

6Q-9

白井 諭* 横尾昭男* 池原 悟* 井上浩子**

(shirai@nttkb.ntt.jp)

*NTT情報通信網研究所

**NTTアドバンステクノロジー

1 はじめに

解析ベースの機械翻訳の精度の向上には、入力文に使われている単語の意味を同定する意味解析（語義解析）の精度の向上が必要である。筆者らは、日英翻訳を対象に、名詞を約3千に分類し木構造の形式に整理した意味属性体系を提案し、主として名詞の意味解析のための単語意味辞書（約37万語）と述語中心のまとまりの意味解析のための構文意味辞書（約1.3万文型）の構築を進めている^[1]。数年間の試行により相互の最適化を図ってきたが、構文意味辞書の強化が重点課題となっている。

最近、黒橋らは例文とソーラスを用いて文型を同定する方法を提案し、精度は解析ベースによる方法と同等以上である可能性を示している^[2]。この方法では、例文の追加により簡単にシステムが改良できる反面、多数の例文を収集する必要や、走行時に多数の例文と照合する必要があることが問題と考えられ、多義性の強い用言（動詞・形容詞と、いわゆる形容動詞）への適用は不利と考えられる。

また、アルモアラムらは自動学習の手法を用いた翻訳ルールの自動抽出方法を提案し、6動詞に対し各27～80の対訳用例を用いた抽出実験に成功しており^[3]、今後、構文意味辞書を改良していく上で考慮しておくべきトピックと言える。

これらの点を踏まえて、本稿では構文意味辞書の現状を分析し、改良の方針について考察する。

2 現状と改良方針

構文意味辞書には、日英翻訳の立場から、日本語の用言に対する英訳語が決定される範囲の格要素を、訳し分けのできる最小限の深さの意味属性を用いて文型として記述し^[1]、日本語の各文型には英語の動詞訳語を中心とした基本構造を対応づけている。また、これらの文型は、一般文型（約6千用言に対し約1万文型）と慣用文型（約3千文型、慣用性を

特徴づける名詞や認定条件、格要素順の制約なども記述）に分類される。本稿で扱う一般文型について、用言と用言あたりの文型数の関係を図1に示す。

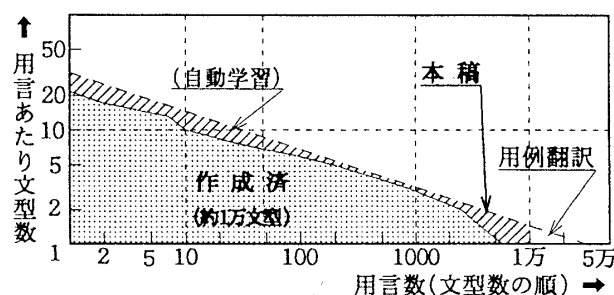


図1 文型の種類とその分布
(文献[1]の一般文型の1993年末時点の合計数)

国際調達仕様書(1361文)を調査したところ、使用されたのは142用言に対する154文型で、うち22用言(22文型)には文型の登録がなく、また23用言に対する25文型が未登録であることがわかった。すなわち、文型数で31%(22+25/154)、用言数で15%(22/142)が不足していることになる。単語意味辞書に約2万用言が収録されていることや各文型の使用頻度の多寡、調査対象の特殊性を考慮すると、構文意味辞書としては最大では図1の破線より下の領域をカバーする必要がある。

ところで、用例ベースの翻訳手法は、一義性の用言への適用は有利であると考えられる。そこで、解析ベースの手法と用例ベースの手法の棲み分けを想定して、当面は用言数1万（文型数2万）を構文意味辞書の守備範囲とすることにする。

一方、自動学習の手法により構文意味辞書の構築に要する工数の大幅削減が期待される^[3]。この手法を適用するには1文型あたり30例以上の対訳用例が必要となるので、現実的には用言数1千程度が適用領域と考えられ、作成済み文型の過不足の検証が中心になると思われる。そこで、本稿では図1の斜線領域のうち右寄りの部分の充足を考える。

3 記述条件の拡張

これまで、和英辞書を参考にして文型の収集を行ない、構文意味辞書を構築してきた。従って、比較的使用頻度の高い文型の収集はほぼ完了していると考えられる。今後は主に対訳用例から文型を収集する予定であるが、1文型あたりの用例数が数件以下に留まることが予想されるため、格要素の意味属性の許容範囲を規定する専門家（アナリスト）の検討の負担が増大すると思われる。

これを軽減するため、格要素の条件として意味属性のほかに単語表記そのものを許容することを提案する。そして、次の手順で文型の収集を進める。

- ①用例の格要素に用いられた単語は、単語表記を格要素の条件として文型を記述する。
- ②類似の用例が出現すれば、作成済みの文型に対して格要素の条件に単語表記を追加する。
- ③単語表記指定が一定個数を超過した文型に対して、意味属性指定への変更を検討する。

例えば、「彼は電話を引いた。→ He installed a telephone.」という用例に対して、①として「X(主体)が/電話”を/引く → X install a telephone」という文型を作成し、その後、日本語動詞が“引く”、英語動詞が“install”である用例が見つかれば②に従ってヨ格の条件を追加し、機会を見て③を行なう。

なお、「名詞+だ(です)」型の日本語の述語名詞は、一般に英語の名詞補語に訳出されるが、「今日は晴れた。→ It is fine today.」や「あなたに質問です。→ I ask you a question.」のように述語名詞が英語の名詞補語には訳出できない場合がある。述語が複合語の場合(XはY次第だ → X depend on Y)もある。日英翻訳の立場からは、これらも構文意味辞書に登録することを併せて提案する。

4 作成支援

前節で述べた文型の収集手順に則って、アナリストによる文型作成作業の支援を考える。

①のステップとしては、日本語の格要素の条件指定と英語の基本構造の生成がポイントである。日本語側は結合価形式であるため、日英翻訳システムの構文解析結果が利用できる。英語側には特定の形式が多用されることがわかっている^[4]。そこで次のような作業支援を提案する(3.~6.を自動化)。

1. 用例の日本語を日英翻訳システムに入力し、訳文と用例の英語の比較から作業対象を抽出する。
2. 用例の英語に基づき、動詞の表記、主語・目的

語等の変数表示を用いて、文型の英語側に記述する範囲を決定する。

3. 英語の基本構造を生成する。このとき文献[4]の情報を参考にする。
4. 主語・目的語等の変数に該当する英単語を、対訳用例の英語から抽出する。
5. 日英対訳辞書により、主語・目的語等の各変数に対応する日本語の格要素を推定する。
6. 格要素の条件には単語表記をそのまま使用することを原則とするが、一部の名詞は意味属性に変更する(“彼”→(主体)など)。
7. 生成結果を確認し、必要に応じて修正した後、構文意味辞書に登録する。

②のステップとしては、既存の文型に条件を追加するだけであるので容易に実現できる。

③のステップとしては、日本語の格要素の条件を単語表記指定から意味属性指定へスムーズに変更できるかがポイントである。もちろん、すべての単語表記を意味属性に置換する必要はないので、次のような観点で意味属性への移行を支援する。

1. 単語意味辞書から表記指定された単語の意味属性を抽出し、共通の意味属性を表示する。
2. 意味属性の木構造の上位で包括される場合、上位の意味属性を表示する。
3. 意味属性の木構造上の上下移動により指定の変更を行なう。

5 おわりに

本稿では、解析ベースの日英翻訳システムの訳文品質に大きく影響する構文意味辞書の現状を分析し、改良方針を示すとともに、改良のための作業手順・作業支援について検討した。今後は、作成中の支援システムを完成させ、用言数1万・文型数2万の構文意味辞書の完成を目指す予定である。

<謝辞>

ご討論くださった上田洋美氏を始めとするNTTアドバンステクノロジーの各位に感謝する。

<参考文献>

- [1] 池原, 宮崎, 横尾: 日英機械翻訳のための意味解析用の知識とその分解能, 情処論 Vol.34 No.8, 1993
- [2] 黒橋, 長尾: 格フレーム選択における意味マーカと例文の有効性について, 情処研報 NL-91-11, 1992
- [3] アルモアリム, 山崎, 秋葉, 横尾, 金田: 概念獲得アルゴリズムの適用による意味属性体系を用いた翻訳ルールの学習, 7回A I 全大 15-3, 1993
- [4] 横尾, 中岩, 白井, 池原: 日英機械翻訳用カルトン-フレッシュ型構文意味辞書の構成, 48回情処全大 6Q-8, 1994