特許文翻訳におけるフレーズアラインメントの評価法

ルールベース翻訳と統計翻訳の融合に向けて

株式会社富士通研究所
ソフトウェア&ソリューション研究所主管研究員
潮田　明

PROFILE
1983年東京大学理学部物理学科卒業。同年株式会社富士通研究所入社。表面磁気光学効果、空間光変調器、統計自然言語処理の研究などに従事。博士（理学）
2007年度から特許版産業日本語委員会委員

ushioda@jp.fujitsu.com

1 はじめに

日英・英日機械翻訳をはじめ多くの実用的機械翻訳システムでは今なお辞書や文法規則に基づくいわゆるルールベース機械翻訳技術が使われているが、90年代から徐々にヨーロッパ言語間での開発が活発化してきた統計翻訳（Statistical Machine Translation、以下SMT）は、近年著しい発展をとげ、最近ではヨーロッパ言語間のみならずアラビア語英語間や中英間などの翻訳においても実用化を目指した開発が進められている。特に非文法的な文の多い話し言葉を扱う音声翻訳においてはSMTは他の翻訳手法に対して圧倒的な強みを発揮している。またテキスト（書き言葉）翻訳においても、大量コーパスを基にした強力な言語モデルを備えていればルールベース機械翻訳よりも格段に自然な訳文を生成することが可能である。

SMTは上記以外にも対訳データさえあれば人手による辞書作成やルール作成が不要というメリットがある反面、大量の学習用対訳データを必要とし、また学習用対訳データとは全く違う分野や文種の翻訳は苦手であるという欠点も併せ持つ。また、ルールベース機械翻訳技術とは異なり、長い文の構造を適切に解析する機構をSMT自体は持たないため、日英間翻訳のように語順や文法が極端に異なる言語間の翻訳においては、特に長い文の翻訳の場合文構造の乱れが著しい。従って、SMTにとって特許文の日英翻訳は今なおハードルの高いタスクであると言える。

そこで最近では、SMTの枠組みの中に、ルールベース機械翻訳等において従来から使われてきた言語学的知識や手法を取り入れる試みが盛んに行われるようになってきた。特に言語構造の大きく異なる日本語と英語の間の機械翻訳においては、SMTにルールベース翻訳や用例翻訳を融合したいわゆるハイブリッド翻訳が有望と思われる。

本稿では、日英間のハイブリッド翻訳において、融合に際してのカギとなるフレーズアラインメントの評価法と評価結果について概説する。

2 フレーズアラインメント

ハイブリッド翻訳において、従来のルールベース翻訳や用例翻訳の資産を最大限に活用しようと考えたときに必ず問題となる重要なポイントは、ルールベース翻訳におけるフレーズとフレーズベース統計翻訳におけるフレーズの間に整合性が全くないことである。前者は構文木におけるconstituentをフレーズの単位として解析を進めるのに対して、後者のフレーズは、言語学的意味とは無縁に、統計的にある意味で有意な単語の連なりをフレーズとして活用している。ハイブリッド方式の中で従来のルールベース翻訳の開発過程で築かれた文法規則あるいは文法記述の枠組みや、言い換え可能性を基準に構築された用例翻訳用の用例の蓄積を有効に活用しようと考えたときには、効率よくconstituentあるいはそれに準拠した対訳フレーズと統計翻訳の枠組みとを組み合わせ

る手立てを考える必要がある。

筆者らは、ハイブリッド翻訳のためのフレーズアラインメントの新しい方式を提案し、その概要をJapio 2007 YEARBOOKにおいて紹介した。詳細は参考文献を参照されたいが、本手法の特長は、統計情報と辞書情報を組み合わせながらボトムアップにバイリンガルパージングを進めることにより統計的最適化の枠組みの中に言語学的制約を組み込むことが可能なことである。従来のフレーズアラインメント手法では、まず両言語間の単語の対応を統計的に同定し、単語対応をもとに徐々に単語列（フレーズ）の対応へとアラインメントを拡張して行くが、拡張の判断基準は基本的に、隣接する単語同士が対応していればその隣接単語（すなわち単語列）も「フレーズ」として対応している、という極めて単純なものであり、単語間の対応の度合い（翻訳確率）や、フレーズの言語学的妥当性と言った要素への考慮はなされていなかった。本手法では、フレーズの拡張の際に、それぞれの拡張が統計的にどれくらい妥当かという統計情報と、拡張されたフレーズが構文的に妥当なものかという言語学的情報を加味することにより、よりハイブリッド翻訳に適したフレーズアラインメントを実現している。

3 フレーズアラインメントの評価

本手法で得られたフレーズアラインメントの評価法とその結果について述べる。フレーズアラインメントを評価する際、フレーズ同士がどれくらい正確に対応付けられているかというアラインメント自体の精度評価と、得られたアラインメントが機械翻訳の精度向上にどの程度寄与しているかという間接的な評価が可能であるが、ここでは前者について考える。単語対応の評価については、その手法も比較的簡単であり、これまでも多くの評価結果が報告されているが、フレーズアラインメントの精度の直接評価についてはまだ定着した手法はなく、報告も少ない。フレーズアラインメントの評価が難しいのは、そもそもフレーズとは何かという本質的問題が絡んでいるからである。（1）を例に取り考えてみる。

（1a）汚水の反応槽滞留時間を短くすることができ、耐久性やコストの面でも満足できる窒素除去装置を提供する

（1b）To provide a nitrogen removing apparatus which can reduce the retention time in a wastewater reaction tank and is satisfactory in terms of durability and costs

Constituentに準拠したフレーズアラインメントでは、対応するフレーズペアの日本語側か英語側の少なくとも一方がconstituentである必要があるが、それ以外にたとえばフレーズの長さなどについては他のアラインメント手法同様、計算の便宜上余り長過ぎないということ以外特別の制約がない。従って、例文（1）のフレーズアラインメントの正解として、以下の（2）、（3）を含む多数の解が存在する。

（2）窒素除去装置を提供する／To provide a nitrogen removing apparatus；
汚水の反応槽滞留時間を短くすることができ、耐久性やコストの面でも満足できる／which can reduce the retention time in a wastewater reaction tank and is satisfactory in terms of durability and costs

（3）窒素除去装置を提供する／To provide a nitrogen removing apparatus；
汚水の反応槽滞留時間を短くすることができ、／which can reduce the retention time in a wastewater reaction tank；
耐久性やコストの面でも満足できる／and is satisfactory in terms of durability and costs

このように正解が複数ある場合、単語アラインメント評価のときのように通常の適合率、再現率のような評価指標をそのまま当てはめるのは難しい。しかし、a）正解が複数あるのならば、システム出力がその内のどれかと一致したなら正解と見なせる b）アラインメントの精度

評価の主な目的の1つが、システムの向上度合いを定量的に測ることにある、という2点から、以下のような評価手法が有効であると考えられる。

1）テスト文として与えられた対訳文について、人手ですべての可能なフレーズ対応を抽出し、それらを正解セット（golden standard）とする。

2）正解セット中のそれぞれのアラインメントに対して、システムが出力したフレーズアラインメント結果についてのF-measure（適合率と再現率のある種の平均値）を計算し、最も高いF-measureをその出力の精度とする。

図1．に2）における適合率と再現率の計算方法を示す。背景の灰色の領域が人手による正解を、太い点線で囲まれた領域がシステム出力の示すフレーズアラインメント結果を示す。領域の面積は升目の数で表す。

日本出願特許の抄録文の課題の部分とその英語訳から160文の対訳文を抽出し、それらに人手で正解を用意して、システムの精度評価を行ったところ、F-meas-

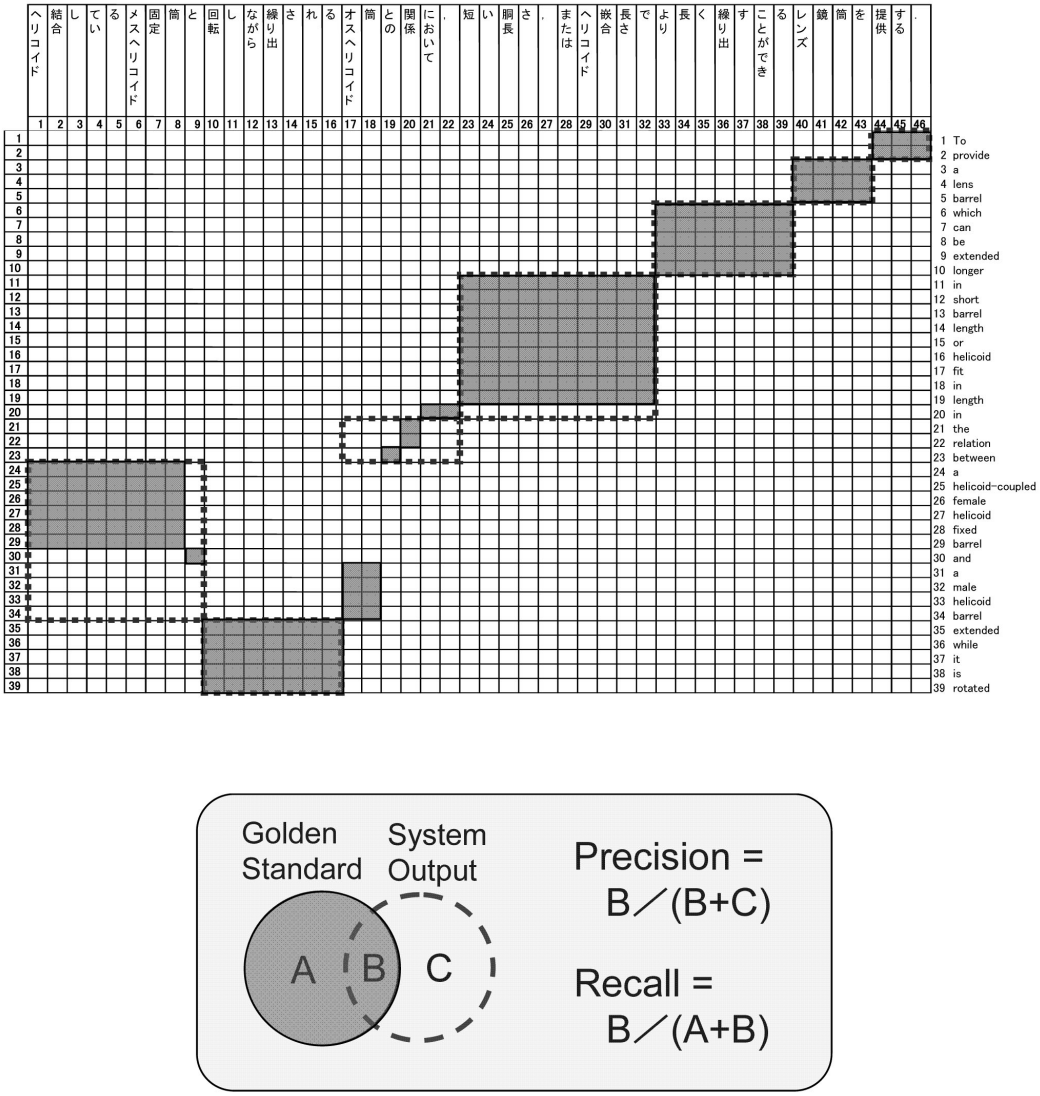

図1　フレーズアラインメントの評価法

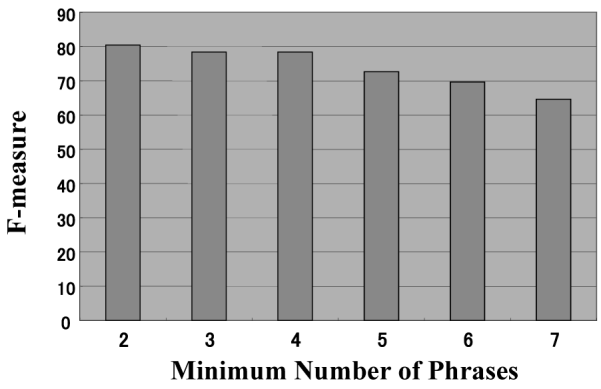

図2　フレーズアラインメントの精度

ureの平均値として80.4という高い精度が得られた。また、各テスト文とシステム出力に対して、最も高いF-measureを与える正解アラインメントの1文中のconstituentの数の平均は6.0であった。このことから、1文を2〜3の大きなフレーズに分割するようなある意味で「安易」なフレーズアラインメントによって高い精度が生み出されているのではないと結論できる。正解セットから、1文中2フレーズ、3フレーズ、…などの粗い分割を順次取り除いていったときのF-measureの変化を図2に示す。図の横軸は正解セットの中での、1文中のフレーズ数の最小値を示す。フレーズ数4までは精度に殆ど影響を及ぼしていないことからも、上の結論が裏付けられる。

4 まとめと今後

ハイブリッド翻訳において重要な鍵を握ると思われるフレーズアラインメントについてその一評価手法と評価結果を紹介した。今後は得られたフレーズアラインメントを基にSMTとルールベース翻訳の融合を図るとともに、本手法による評価と翻訳精度との関連を調べて行きたい。

【参考資料】

The TC-STAR project http://www.tc-star.org/

潮田明「ハイブリッド翻訳のためのフレーズアラインメント」Japio 2007 YEARBOOK, 2007.

Akira Ushioda（2007）“Phrase Alignment Based on Bilingual Parsing.”Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation, pp.241-250.

Melamed, I. Dan（1997）. A Word-to-Word Model of Trans-lational Equivalence. In Proceedings of the Eighth Con-ference of the European Chapter of the Association for Computational Linguistics（pp.490-497）.

Franz-Josef Och and Hermann Ney（2004）“The alignment template approach to statistical machine translation.” Computational Linguistics, 30（4）, pp.417-450.

Kenji Yamada and Kevin Knight（2001）“A syntax-based statistical translation model.” Proceedings of the 39th Annual Meeting of the ACL, pp.523-530.