

制約付き有向グラフを利用した文生成

5R-7

武田 浩一

日本アイ・ビー・エム株式会社 東京基礎研究所

1 まえがき

機械翻訳における文生成では、発話あるいは文そのものには含まれない情報が、適切な翻訳のために有効に利用できる。従来の中間言語表現を用いた翻訳手法では、文単位に中間言語で表現された文の意味を、すべて目的言語で表現するという方法をとることが多かったため、文脈情報や目的言語の表層表現を作成する際に課される制約のもとで、最適な翻訳を実現することができなかった。本論文では、文脈や表層表現の与える制約情報と、翻訳したい文の内容を同時に表現できるような中間言語表現を提案し、このような中間言語表現からの文生成アルゴリズムについて述べる。

2 制約付き有向グラフ

文のもつ意味と、種々の制約を同時に表現するための道具として、制約付き有向グラフ [6] を用いる。制約付き有向グラフは、通常の素性構造を表現する有向グラフの各節点および各有向枝に k 個の「色」という属性を追加したものである。簡単のため、ここでは $k=3$ とし、赤、黄、緑、の3色を考える。色は、有向グラフの要素に対して、重要度を定義するために用い、各色の直観的な意味は次の通りとする。

- 赤色の節点あるいは有向枝は、それが必須の情報であることを示す。
- 黄色の節点あるいは有向枝は、それと矛盾する情報を許さないことを示す。
- 緑色の節点あるいは有向枝は、それと矛盾する情報があってもかまわないことを示す。

従って、上記の色をもつ有向グラフを文の中間表現とした場合には、赤色の要素をすべて含み、黄色の要素と矛盾しないような中間表現の集合が、目的言語として許される文の候補を定めるものと解釈できる。従来の知識ベース翻訳システム [2] での中間表現は、すべての節点と有向枝が赤色である場合に相当する。

節点と有向枝の重要度が一貫性を持つように、任意の制約付き有向グラフ G に以下のような整合性の条件を仮定する。

● G の根は赤色である。

- 赤色の有向枝は、赤色の節点どうしを接続する。
- 任意の赤色の節点は、根から赤色の有向枝のみによって到達可能である。
- 黄色の有向枝は、赤色または黄色の節点を接続する。
- 任意の黄色の節点は、根から赤色または黄色の有向枝によって到達可能である。
- 黄色の節点あるいは有向枝は、それと矛盾する情報を許さないことを示す。
- 2つの節点の間に同じラベルで色の異なる有向枝があってはならない。

制約付き有向グラフの包摂関係とは、通常の有向グラフの包摂関係 [5] に以下のような拡張を含めたものとする。

- 赤色の節点(有向枝) は、赤色の節点(有向枝) の包摂を包摂する。
- 黄色の節点(有向枝) は、黄色または赤色の節点(有向枝) を包摂する。
- 緑色の節点(有向枝) は、任意の色の節点(有向枝) を包摂する。
- 2つの制約付き有向グラフの単一化についても、同様に定義できる。この際の色に関する拡張は、
 - 赤色の節点(有向枝) は、任意の色の節点(有向枝) との単一化によっても赤色である。
 - 黄色の節点(有向枝) は、黄色または緑色の節点(有向枝) との単一化のもとで黄色である。
 - 2つの緑色の節点(有向枝) の単一化は、緑色の節点(有向枝) である。

となる。通常の単一化では、2つの異なる値を表現する節点どうしの単一化は失敗するが、上記の制約付き有向グラフでは、一方が緑色の節点の場合には他方が、その節点の値を上書きするという解釈を許すことができる。これにより、いわゆる defeasible reasoning のように反

駁可能な仮説を立てて、文の意味を解析していくという手法の中間言語としても制約付き有向グラフが利用できることがわかる。矛盾する2つの緑色の節点を単一化する場合は、2つの対立する仮説が存在すると考えて、グレード付き単一化[3]を適用したり、その節点の単一化を遅延するといった方法がとれる。

3 概念トランスファー規則

Dorr[1]がdivergenceとして整理しているように、原言語と目的言語での中間言語表現上のギャップは多様であり、これらの異なる表現の対を単純にトランスファー規則で対応づけるのは、トランスファー規則のコーディングに多大な労力を要するのみでなく、中間言語方式の利点である、多言語翻訳を非実用的にになってしまうことになる。この問題を解決するために、前節で述べた制約付き有向グラフを用いて、原言語の中間言語表現から、段階的に少しでもましな翻訳を可能にするようにできる。このような方式では、概念トランスファー規則の主な役割は、原言語の中間表現に含まれる情報をできるだけ失わないように、制約として目的言語の中間表現に残しつつ、翻訳可能な中間表現に変換することになる。例えば、英語で、直接目的語と不定詞を補語にとる動詞wantの用法の中間言語表現に対する日本語生成

```
(*want-2
(:agent (*X))
(:theme (*Y))
(:goal (*Z)))
```

(*X)は、(*Z)に(*Y)を望む

への概念トランスファー規則(ej)として、

```
(ej *want-2
((:theme)=yellow(:goal:agent)))
```

で、もとの概念表現の*want-2の:theme フィラーを、:goalの:agent フィラーと同一であるという制約(=yellow)を追加する。これにより、日本語生成時に「～を」の行為の意味上の主語が制約として概念表現である制約付き有向グラフでの黄色の節点として表現され、行為の適切な訳語選択に寄与することができる。もし、原言語の解析時に、このような不定詞の制御(control)が文法で処理される場合には、既に存在する:goalの:agent フィラー(赤色の節点)を黄色の節点に弱めるような規則を用いるとよい。こうすることで、日本語生成時に「(*Z)に(*Z)が～すること」を望む」という具合に動作主が2回翻訳されることを防げる。

文脈処理[4]を採用している場合には、さらに代名詞の参照決定や以前に言及された内容の繰り返しの検出により、中間言語表現の特定の節点や有向枝の色を変化させることができ、訳文を改善することができる。

4 文生成アルゴリズム

以上のような道具立てにより、文生成アルゴリズムは次のようなステップで構成できる。今、入力となる制約付き有向グラフをGとする。

1. 文生成モジュールは、概念・表層表現の写像規則[7]を用いて、Gのすべての節点と有向枝に対応する表層表現を合成する。
2. Gのすべての赤色の節点と有向枝の部分が表層構造に写像できた場合には、それがGの目的言語への翻訳になる。
3. Gのある節点を根とする部分グラフが表層構造に写像できない赤色の節点あるいは有向枝を含む場合には、概念トランスファー規則を適用し、可能な変換を調べる。
4. 適切な概念トランスファー規則が存在しない場合には、問題となる節点あるいは有向枝を黄色に変える。
5. それでも表層構造の生成に失敗する場合には、問題となる黄色の節点あるいは有向枝を緑色に変える。

このようなアルゴリズムのもとで生成される文の表層構造は、過大生成(over-generation)や過小生成(under-generation)のない、もとの入力Gの赤色の節点や有向枝を含み、Gの黄色や緑色の節点や有向枝の一部を含むような範囲に対応することが示せる。

参考文献

- [1] B. J. Dorr. "Machine Translation: A View from the Lexicon". The MIT Press, Cambridge, Mass., 1993.
- [2] K. Goodman and S. Nirenburg, editors. "The KBMT Project: A Case Study in Knowledge-Based Machine Translation". Morgan Kaufmann Publishers, San Mateo, California, 1991.
- [3] A. Kim. "Graded Unification: a Framework for Interactive Processing". In *Proc. of the 32nd Annual Meeting of ACL*, pages 313-315, Las Cruces, New Mexico, June 1991.
- [4] T. Nasukawa. "Robust Method of Pronoun Resolution Using Full-text Information". In *Proc. of the 15th International Conference on Computational Linguistics*, pages 1157-1163, Kyoto, Japan, Aug. 1991.
- [5] C. Pollard and I. A. Sag. "An Information-Based Syntax and Semantics, Vol.1 Fundamentals". CSLI Lecture Notes, Number 13, 1987.
- [6] K. Takeda. "Tricolor DAGs for Machine Translation". In *Proc. of the 32nd Annual Meeting of ACL*, pages 226-233, Las Cruces, New Mexico, June 1991.
- [7] K. Takeda, N. Uramoto, T. Nasukawa, and T. Tsutsumi. "Shalt2 - A Symmetric Machine Translation System with Conceptual Transfer". In *Proc. of the 14th International Conference on Computational Linguistics*, pages 1034-1038, July 1992.