

日独音声言語翻訳実験システム

6B-6

鈴木 雅実 菊井 玄一郎 Mark Seligman Herbert Tropf* 森元 逞 樽松 明
ATR 自動翻訳電話研究所

1 はじめに

これまでATRでは、日英音声言語翻訳実験システムを中心に研究開発を進めてきたが、海外の研究機関との共同による、3ヶ国間の双方向音声翻訳通信プロジェクト(CSTAR)の発足に伴い、ドイツ語翻訳への拡張を実施することとなった。日独版の作成に際しては、日英版への使用を前提とした翻訳処理システムが、目標言語の違いに対して柔軟に適応可能であることを実際に検証することも、一つの目的として開発を行なった。その結果、現在までの比較的短期間に所期の目標を十分達成したので、以下では日英版と適宜比較しながら、この日独システムの現状を報告し、今後の展望と課題を述べる。

2 音声言語翻訳システム日独版への拡張

ATRの音声言語翻訳実験システムの設計や技術的な課題については参考文献[5][7][4]に詳しいが、ここでは言語対の拡大に際する問題提起を通して、各処理部で採られた方針等について説明する。

2.1 開発手順

日英版として開発した実験システムを日独版へ拡張するのに先だて、後述する対象範囲について想定翻訳結果を作成し、特に語彙・構文面からの分析を行なった。これは、日本語・英語・ドイツ語を比較する形で行なわれた。この段階で、同一表現に対して一貫した対訳をあてることを原則としたが、冠詞や数の生成を含む多義については、今回の開発段階での本格的な処理を留保し、翻訳結果には主に領域内での頻度によるデフォルト値が反映されることとなった。

次に、ドイツ語生成の入力となるドイツ語素性構造の仕様を、英語版での経験に基づいて設計した。この作業は、言語変換処理および生成処理の双方に関わるインタフェースの決定を意味するため、多くの時間を費やした。この素性構造の仕様を下敷に、ドイツ語生成文法の作成するとともに、この構造を得るための日独変換規則を作成した。実際には一定の試行期間の後に、再度意味構造の見直しを行なった。また、作成済みの日英変換規則の再利用がある程度可能であった。

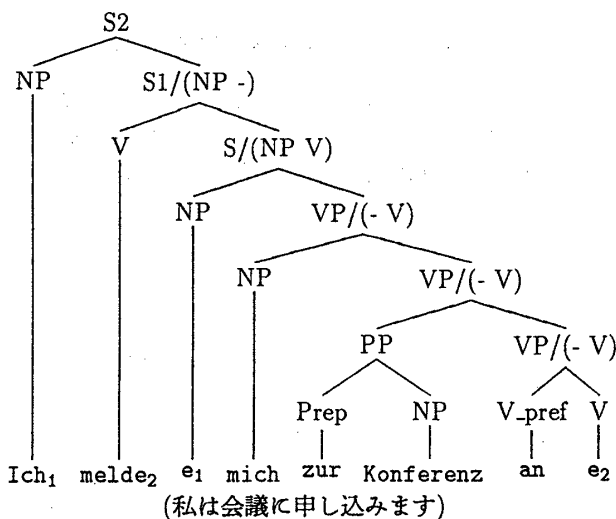
2.2 ドイツ語生成

日本語から見て、英語とドイツ語はともに西欧言語に属し、互いに距離の近い言語同士である。しかし、いうまでもなく、各々の文法体系は他方ない特徴を抱えており、その固有性をいかに効率良く記述するかということが、生成文法を記述する上で重要である。

たとえば、ドイツ語では動詞の基本的な位置として、文頭、2番目、文末という3つの位置が考えられ、これらの位置に応じて動詞の表層形や(いわゆる分離動詞における)接頭辞の分離の有無などが制約される。このような文法的制約を1)スラッシュ素性を用いた動詞位置の制約、2)形態的素性を用いた表層形の制約、などを用いて語彙主導型句構造文法の枠組で一般的に記述した。このほか、再帰動詞に出現する再帰形代名詞に関する情報は動詞の語彙から供給される。[8]

生成文法は基本的に単一化に基づく句構造文法の形式であるが、同一語幹の語に関しては屈折形ごとに語彙エントリを設けるのではなく、1エントリとして扱い、句構造生成後にdecision netを文法として用いた処理系で屈折変形を行なっている。これらの処理は英語と同一の処理系で実現した。[2][3]また、ドイツ語では格変化など英語と比較して複雑な形態素生成処理が必要である。形態素生成処理は句構造生成と協調して適切な語尾変化を行なっている。

以下にドイツ語の木構造と生成結果の例を示す。



注) e1, e2 は共に空範疇でそれぞれ Ich1, melde2 に対応する

図1. ドイツ語の木構造と生成結果の例

An Experimental Japanese-German Speech Translation System
Masami SUZUKI, Gen-ichiro KIKUI, M. SELIGMAN,
H. TROPF, Tsuyoshi MORIMOTO, Akira KUREMATSU
ATR Interpreting Telephony Research Laboratories
*Exchange researcher from Siemens AG., Germany

2.3 日独変換

一方、言語変換処理の観点から見ると、同一の入力日本語文の構造に対して、常に英語・ドイツ語で同様の(相似的な)構造を対応させることが最適であるとは限らない。NativeやBilingualの話者が作成した、期待される(好ましい)翻訳結果を見ると、意味的・統語的に同型の構造が可能であっても、実際には大きく構造が異なっている場合もある。これには次のような背景が考えられる。

1. 慣用的に好まれる構造が英・独語間で異なる場合
2. 直訳的な表現の許容度が英・独語間で異なる場合

上記の二点のうち1の典型的な例の一つは、多くの場合日本語で主語が明確に表現されていない文で、英語なりドイツ語で、次のいずれかが好まれるという問題である。一意に決定することが困難な場合もある。

- 1) 人の主語を立てた文
- 2) 非人称構文
- 3) 受身文

これは使用される語彙や慣用に影響される(下記の翻訳例1を参照)。本システムでは日本語素性構造から目標言語の素性構造への写像を行なうトランスファー方式を採用しており、意図表現以外の命題部分については構造中で語彙の選択が行なわれるため、これらの選択は変換処理が受け持つことが多い。

2については出現頻度等のデータを持ち合わせていないが、一例として「手続きをする」という日本語表現に対する翻訳を考える(下記の翻訳例2を参照)。E-2では、入力文の構造に沿った直訳としているのに対し、G-2では、文全体の構造を捉えて構造を変化させている(このドイツ語文を直訳すると、どのように先へ進めばよいか、ということになる)。これは英語と同様の直訳的な構造の文では、ドイツ語としての許容度(自然性)が極めて低いためである。元の日本語文はいわゆる慣用句とも異なるので、その取り扱いが難しい。

日独・日英の翻訳実験結果の例

以下に示すのは翻訳実験結果のごく一部である。J-, G-, E- で始まる記号を付した文はそれぞれ、入力日本語文・ドイツ語翻訳結果・英語翻訳結果を表す。

J-1 九月二十七日以後の取消しに対する払い戻しはできません

G-1 Eine Zurückzahlung für eine Abmeldung nach dem siebenundzwanzigsten September ist nicht möglich.

E-1 I can not make a refund to the cancellation after September 27th.

J-2-prev 会議に参加したいのですが(直前の発話)

J-2 どのような手続きをすればよろしいでしょうか

G-2 Wie soll ich vorgehen?

E-2 What kind of procedure should I follow?

また、変換処理の手法自体についての詳細は他の文献[1]等に譲るが、日英システムと同様、入力日本語素性構造からの逐次的な構造書き換え処理により、最終的なドイツ語素性構造を得ている。日独版での改善点は、否定要素の移動等の特定の構造変換用に、書き換え処理のサブ過程をやや詳細化し、規則の適用効率の向上を図った点である。

3 処理の現状と今後の課題

日独翻訳の対象として扱ったのは、国際会議に関する問合せをタスクとしたモデル会話12ファイルで、約230の異なり文からなるコーパスである(語彙数にして約700語の規模)。対象文の翻訳結果は概ね満足できる結果となっている。日英版の方は、約1500語でより多様な構造を扱うべく、拡張を進めてきており、一般性および頑強性の点で日独版はこれに及ばない。また、日英版で試行中の文脈情報や領域知識を用いた翻訳処理の本格的導入は今後の課題である。[6]

これまで述べたように、本システムの扱える範囲は現在のところ限定されている。今後、量的質的な拡大や、さらに他の言語への拡張を考えるとときに必須となるのは、個々の言語の多様な言語表現を素性構造上でいかに合理的かつ簡潔に表現するか、という問題である。これは変換処理と生成処理のインタフェースの設計の観点からも、文脈や対象領域等の外部知識を効率よく参照するためにも重要である。

4 まとめ

ATRの音声言語翻訳実験システムによる日独翻訳の実践について報告した。対話文という従来の経験の蓄積が乏しい対象について、複数の目標言語への翻訳という観点から、ほぼ並行的に翻訳処理を展望できたことは非常に有益であった。処理の現状については、基礎固めの段階で今後なすべき課題は多いと言える。紙面の構成上割愛せざるを得なかった、個別の興味深い問題についての検討は別の機会に譲りたい。

参考文献

- [1] 長谷川 敏郎: 素性構造書き換えシステムのSL-TRANS変換過程への適用, 人工知能学会全国大会, 1990.
- [2] 菊井 玄一郎・鈴木 雅実: 表層表現の型を用いた対話文の生成について, 情報処理学会研究報告, NL-81-11, 1991.
- [3] G. Kikui: Feature Structure Based Semantic Head Driven Generation, *COLING 92*, 1992.
- [4] 樽松 明: 自動翻訳電話のための音声処理と言語処理, 電子情報通信学会誌 Vol.75 No.10, 1992.
- [5] T. Morimoto, M. Suzuki et al.: A Spoken Language Translation System: *SL-TRANS 2*, *COLING 92*, 1992.
- [6] 鈴木 雅実: 対話文の日英構造変換における関連知識の利用～基礎検討～, 電子情報通信学会技術研究報告 NLC92-13, 1992.
- [7] 竹澤 寿幸 他: ATR 音声言語翻訳実験システム ASURA, 情報処理学会第46回全国大会, 6B-5, 1993.
- [8] Herbert S. Troup: The German Grammar for ASURA, *ATR Technical Report*, TR-I-0289, 1992.