

ATR 音声言語翻訳実験システム ASURA *

6B-5

竹沢寿幸 森元 逞 谷戸文廣 鈴木雅実 嵯峨山茂樹 樽松 明†

ATR 自動翻訳電話研究所‡

1 まえがき

異なる言語を話す人同士のコミュニケーションが円滑に行なえることを目指して[1]、音声認識・言語翻訳および音声合成を一貫して結合した音声言語翻訳実験システム ASURA (Advanced Speech Understanding and Rendering System of ATR) を作成した。音声認識部は ATR で提案・開発された SSS-LR 方式[2, 3]を採用しており、多数話者に対しても高い認識率を達成している[4]。言語翻訳部は ATR で提案・開発された話し言葉翻訳手法を採用しており、英語の他に新たにドイツ語も出力するように拡張している[5]。国際会議に関する問合せを実験タスクとし、音声認識部・言語翻訳部ともに約 1,500 語の語彙を扱っている。本報告では、そのシステムの概要と処理性能を述べる。ドイツ語への翻訳についての詳細は文献[5]に譲る。話し言葉翻訳の評価については文献[6]に譲る。

2 システムの概要

発話者は、文節毎にポーズをおいて、文単位で発話する。音声区間はパワーとゼロ交差数を使って自動的に検出される。

言語翻訳における日本語解析部は日英・日独ともに共通に利用している。また、英語への変換・生成とドイツ語への変換・生成の処理系は基本的に同じものである。そこでの規則が異なるだけである。

2.1 システムの特徴

1. 音声認識における音素モデルは、前後の音素の影響を詳細に反映した異音モデルであり、隠れマルコフ網で表現されている[2]。
2. 音素モデルを駆動するものとして、音素環境依存一般化 LR 構文解析法を実装している[3]。
3. 多数話者に対して高い音声認識率を達成するために、移動ベクトル場平滑化方式を提案し、少量の学習データでの話者適応を可能としている[4]。
4. 文節内文法と文節間文法を使って、統語的に妥当な文候補を音声認識スコアの順に出力するので[7]、文節単位の発話であっても、文を単位として高い認識率を達成している。
5. 言語翻訳のための日本語解析部では、日本語の話し言葉に頻出する省略表現や、間接的な依頼のような多種多様な表現を扱うために、統語・意味・語用論的制約を単一化文法 (Unification-Based Grammar) で記述している[8]。
6. 変換部では、処理系と変換規則を完全に分離しているので、他の言語への翻訳系の開発が容易である[9]。その優

れた拡張性を利用して、短期間で新たにドイツ語を出力できるようにできた[5]。

7. 生成部は、単一化文法を基本とした宣言的な規則記述に基づく処理系であり、慣用表現にも柔軟に対応できる句構造規則を提案・実装している[10]。つまり、解析部と生成部は同じような思想で設計されているので、二つの間で文法の共通化を図ることもできる。

2.2 システム構成

ASURA システムの構成を図1に示す。ワークステーション HP9000/750 上で動作する。英語音声合成は DEC Talk を利用し、ドイツ語音声合成はベンツ社の試作品を利用している。

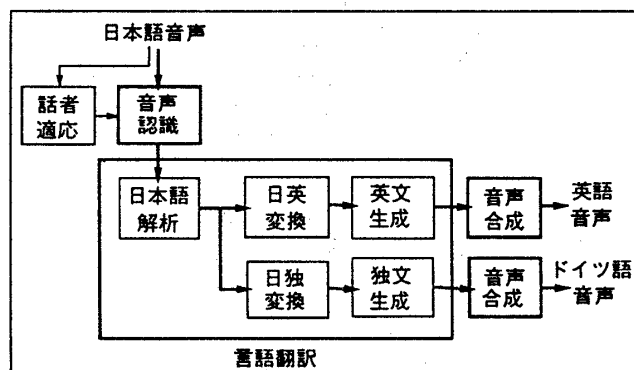


図1: システム構成

3 システムの性能

国際会議に関する問合せを実験タスクとし、音声認識部・言語翻訳部ともに約 1,500 語の語彙を扱っている¹。言語翻訳部では、日本語話し言葉の標準的な言いまわしの約 90% を扱うことができる[6]。システムの翻訳例を表1に示す。入力日本語中の“/”は文節区切りを表している。

いろいろな条件下での音声認識率および英語への翻訳率を表2に示す。評価用の文は平均文節数 2.71 の 259 文セットであり、話者適応は 25 単語セットで行なった。話者名の頭文字に M が付いているのが男性話者、F が付いているのが女性話者である。話者は特殊な訓練を受けていない関東地区出身の一般人である。文節間の処理で予測型というのは文節間文法を音声認識時に利用するもので、ラティス解析というのは文節毎に独立に音声認識を行なった後で文節間文法を使って文候補を組み立てるものである[7]。ビーム幅は音声認識時の探索過程の枝の数の上限値である。表2には結果の一部のみ示している。

文節間文法を予測的に利用した方が、いずれの話者に対しても若干高い認識率が得られた。処理の効率(計算時間)は予測

*ASURA: Advanced Speech Understanding and Rendering System of ATR

†Toshiyuki TAKEZAWA, Tsuyoshi MORIMOTO, Fumihito YATO, Masami SUZUKI, Shigeki SAGAYAMA, Akira KUREMATSU

‡ATR Interpreting Telephony Research Laboratories

¹ただし、ドイツ語への翻訳系は約 700 語の規模。

表 1: システムの翻訳例

入力日本語文	出力英語文	出力ドイツ語文
もしもし。	Hello.	Hallo.
会議に／申し込みたいのですが。	I'd like to apply for the conference.	Ich möchte mich zur Konferenz anmelden.
どのような／手続きを／すれば／よろしいのでしょうか。	What kind of procedure should I follow?	Wie soll ich vorgehen?
登録用紙を／至急／送らせていただきます。	I will send you a registration form immediately.	Ich schicke Ihnen sofort ein Anmeldeformular.

表 2: 音声認識率および英語への翻訳率

話者	文節間の処理	ビーム幅	文節認識率 (%)			文認識率 (%)			英語への翻訳率 (%)
			1位	～2位	～3位	1位	～2位	～3位	
MIK	予測型	250	91.3	96.2	97.2	84.9	90.3	91.1	90.7
	予測型	100	87.6	91.6	92.8	77.2	81.1	81.5	84.2
	ラティス解析	250	91.2	95.9	97.0	84.9	90.0	90.3	90.4
MST	予測型	250	87.2	93.2	95.6	76.4	83.4	86.9	86.1
	予測型	100	84.1	88.9	91.2	70.7	74.5	76.8	77.2
	ラティス解析	250	87.2	93.0	95.9	76.1	83.0	86.1	86.5
FAK	予測型	250	95.3	97.4	98.6	90.7	94.2	95.0	91.1
	予測型	100	92.9	95.0	96.0	85.3	88.4	89.2	86.5
	ラティス解析	250	94.3	97.0	98.3	89.2	92.3	93.4	90.7
FNY	予測型	250	87.9	95.5	96.4	77.6	89.2	91.1	86.1
	予測型	100	85.5	91.9	92.8	74.5	83.0	84.2	81.1
	ラティス解析	250	87.6	96.2	97.3	77.2	88.8	91.1	86.9

型でもラティス解析でもほぼ同じであるので、文節間文法は予測的に利用した方がよいと言える。

英語への翻訳率は、3位までの文候補の範囲で最初に翻訳可能であった文が正しい英語と判定された率である。正解より上位にある意味・語用論的に不適格な文候補を日本語解析部で除去することができるため、英語への翻訳率が1位の文認識率より必ず高い。「会議に申し込みたいのですが」が「会議に申し込みたいのですが」となるような若干の誤認識が含まれていても正しく翻訳できることがあるため、英語への翻訳率が3位までの文認識率より若干高いことさえある。逆に、英語への翻訳率が3位までの文認識率より若干低いこともある。その理由は少なくとも2つある。1つは、正解より上位にある別の意味を持つ文が英語に翻訳可能となることがあるためである。もう1つは、一文単位の翻訳では曖昧性が解消できず適切な訳し分けができないことがあるためである。また、語彙数の増加に伴って同音意義語の存在が無視できなくなってきた。

さらに、より多様な言いまわしを含んだ平均文節数3.87の559文セットについても評価実験を進めている。

4 むすび

音声言語翻訳実験システム ASURA について報告した。言語翻訳部での優れた拡張性を利用して、英語の他に新たにドイツ語も出力するようにした。本報告では、そのシステムの概要と処理性能を述べた。今後は、より自由な発話を許す、高度な音声翻訳の実現を目指す。

謝辞

本システムは、筆者ら以外にも多くの人の努力により完成することができた。鷹見淳一、永井明人氏には話者適応機能付きの SSS-LR 連続音声認識器を利用可能にもらった。保坂順子氏には音声認識のための文法を作成してもらった。Harald Singer 氏には便利な音声入力プログラムを作成してもらった。さらに、性能評価のため

の音声認識実験を手伝ってもらった。田代敏久、永田昌明氏には翻訳のための日本語解析部の拡張をもらった。菊井玄一郎、Herbert S. Tropic, Mark Seligman 氏にはドイツ語への翻訳を新たに作成してもらった。記して感謝する。

参考文献

- [1] T. Morimoto, T. Takezawa, et al: "Enhancement of ATR's Spoken Language Translation System: SL-TRANS2", *Proc. of ICSLP-92*, pp. 397-400 (1992-10).
- [2] 鷹見淳一, 嵯峨山茂樹: "音素コンテキストと時間に関する逐次状態分割による隠れマルコフ網の自動生成", *信学技報*, SP91-88 (1991-12).
- [3] 永井明人, 鷹見淳一, 嵯峨山茂樹: "逐次状態分割法 (SSS) と音素コンテキスト依存 LR パーザを統合した SSS-LR 連続音声認識システム", *信学技報*, SP92-33 (1992-06).
- [4] 鷹見淳一, 永井明人, 嵯峨山茂樹: "逐次状態分割法 (SSS) と LR パーザを統合した SSS-LR 連続音声認識手法における話者適応の性能評価", *日本音響学会平成4年秋研究発表会講演論文集*, 2-5-5 (1992-10).
- [5] 鈴木雅実, 菊井玄一郎, M. Seligman, H. Tropic, 森元逞, 樽松明: "日独音声言語翻訳実験システム", *情報処理学会第46回全国大会*, 6B-6 (1993-03).
- [6] 浦谷則好, 菊井玄一郎, 田代敏久, 田窪行則, 定延利之, 成田一: "話し言葉の日英翻訳システムの評価法", *情報処理学会第46回全国大会*, 6B-4 (1993-03).
- [7] 竹沢寿幸, 保坂順子, 北研二, 森元逞, 江原暉将: "構文規則を用いた文音声認識", *信学技報*, SP90-73/NLC90-45 (1990-12).
- [8] M. Nagata, T. Morimoto: "A Unification-Based Japanese Parser for Speech-to-Speech Translation", *IEICE Trans. Inf. & Syst.*, E76-D, 1 (1993-01).
- [9] T. Iisegawa: "A Rule Application Control Method in a Lexicon-driven Transfer Model of a Dialogue Translation System", *Proc. of ECAI-90*, pp. 336-338 (1990-08).
- [10] G. Kikui: "Feature Structure Based Semantic Head Driven Generation", *Proc. of COLING-92*, pp. 32-38 (1992-08).