

## 音声言語日英翻訳実験システム(SL-TRANS)の概要

4G-4

森元 還 小倉 健太郎 相沢 輝昭 横松 明

ATR自動翻訳電話研究所

### 1. はじめに

言葉の障壁克服の一つとして自動翻訳電話<sup>[1]</sup>実現のための基礎研究を行っている。本稿では、そのための第一歩として音声認識、機械翻訳、音声合成を統合した翻訳システム“SL-TRANS”について述べる。

### 2. システム構成

SL-TRANSは、日本語の話し言葉の音声を英語へ翻訳して音声出力する音声言語翻訳システムであり、三つのサブシステムからなる分散型システムである。一つは音声認識処理を行う部分、もう一つは言語処理をする部分、さらに音声処理と言語処理を結び付けるための統合処理部分である。このような構成は各処理のモジュール性が高く、既存の音声認識処理と言語処理を容易に結合できる。

音声認識処理の結果は非常に多くの候補を残したものになっているので、これを直接言語処理に渡すと言語処理の負荷が重く成り過ぎる。そこで、音声認識処理の結果を簡易な処理によって、音声認識処理の結果を絞り込む必要がある。統合処理では、係り受けの統計的な情報を利用して絞り込みを行っている。

### 3. 処理の流れ

#### (1)文の選択および選択された文の音声出力

全体のコントロールおよび結果の表示は統合処理で行う。統合処理で、まず処理する文の選択やその文の音声出力をを行う。

#### (2)音声認識

SL-TRANSでは、音声認識の簡単化のため、音声データは予め文節ごとに区切って発生したものである。処理する文が選択され、翻訳実行要求を受けると、プロセス間通信機能を用いて音声認

識処理のプロセスを呼び出す。

音声認識では、HMM(Hidden Markov Model)音韻認識<sup>[2]</sup>と予測LRパーザを用いた文節認識を行っている。言語的な知識を文脈自由文法で記述し、解析の効率を向上させるために、それを拡張LR構文解析法における動作表で表現する。文法を音韻予測に使うため、LR文法の終端記号は普通の文法のように単語ではなく、音韻になっている。動作表を用いて音声データ中の音韻を予測し、予測された音韻の存在確率をHMM音韻照合モジュールで調べる。このような方法により、文法情報を参照しながら効率が良く精度の高い音声認識を実現している。音声認識の結果は尤度つきの文節ラティスである。文節ごとに最大5つの候補が出力される。正解の候補がこの段階で落ちないように、ほとんどの場合候補は5つ得られる。

#### (3)係り受け処理と信頼性評価

音声認識結果の信頼性を結果の尤度を用いて評価し、信頼できないと判定した文節に対してはユーザに問合せを行う。(将来的には音声による再入力を行う。)

次に、係り受けによる絞り込みを行い、言語処理に絞り込んだ文節ラティスを渡す。係り受けを利用できるものについては各文毎の候補が高々二

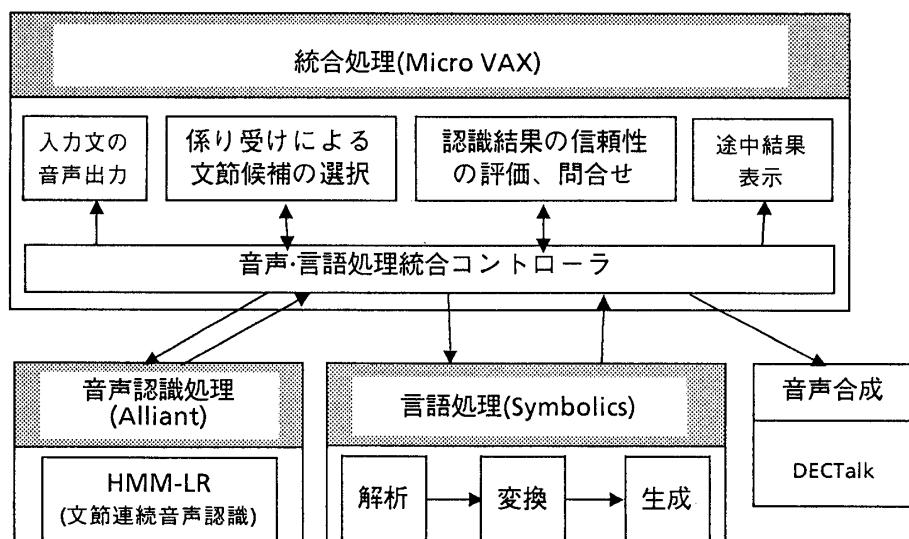


図1. システム構成

つになる程度に文節候補を絞り込む。係り受け処理で係り受け関係が付けられない孤立した文節に対しても、音声認識が信頼できないと判断し、問合せ処理を起動し、正しい文節候補をユーザに選んでもらう。

係り受け関係の係る係らないの判断は、国際会議に関する問合せを題材とした会話の言語データ<sup>[3]</sup>の係り受け情報を利用して行っている。

#### (4) 言語処理

言語処理では、文節ラティスを入力として解析、変換、生成を行い結果として英文を出力する。間接表現や敬語表現が複雑であり、主語などの省略が多いなどの、日本語の話し言葉に特徴的な問題点を、語彙駆動単一化文法を用いた素性構造伝搬解析手法<sup>[4]</sup>をもちいて解決している。また、アクティブ・チャート解析機構で、音声認識結果の尤度等を考慮し解析する。

#### (5) 音声合成

言語処理の結果は音声合成器を使って、英語のテキストから規則による音声合成を行い、音声合成音声出力する。

### 4. マンマシンインタフェイス

図2にSL-TRANSのマンマシンインタフェイス画面の一例を示す。マンマシンインタフェイスはX-Windowを利用して実現している。入力文表示画面、翻訳文表示画面、コマンド入力画面、音声認識結果表示画面、係り受けによる候補選択画面、問合せ画面からなる。入力は総てマウスだけを使って行え良好なマンマシンインタフェイスを実現している。

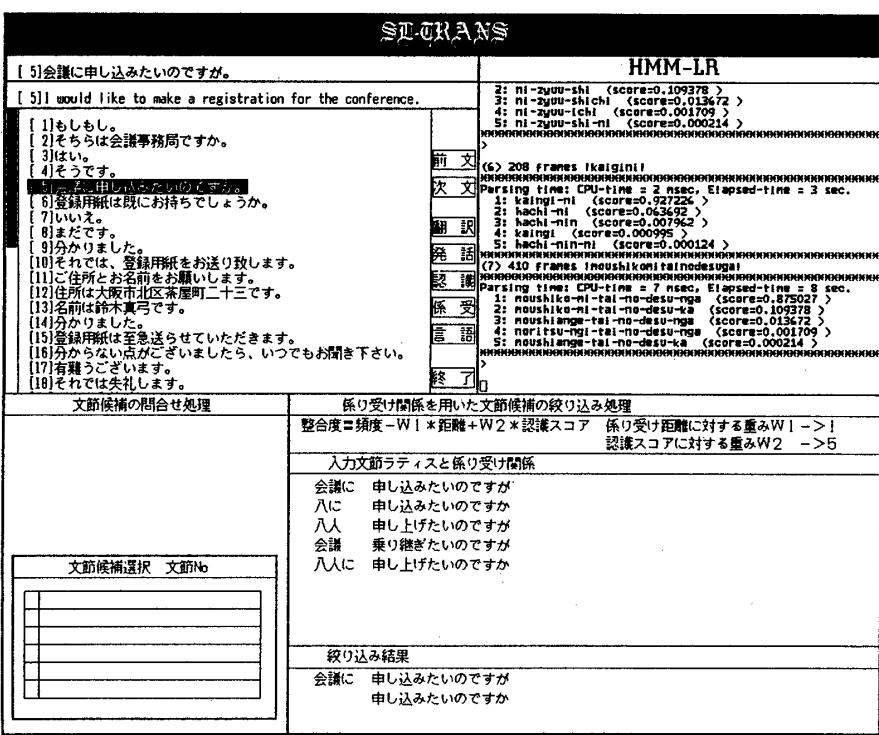


図2. マンマシンインタフェイス

### 5. システムの性能

現在、国際会議に関する問い合わせを題材として翻訳実験を行っている。文法の語彙数300語の条件下で、2会話のモデル会話(37文)に対して翻訳を行った結果、音声認識から翻訳文出力まで、1文あたり約1分、(音声認識10秒、係り受け5秒、言語処理45秒)程度で翻訳できた。音声認識率は、正解が第一位にある率は87%、第五位以内に正解が含まれる率は96%である。音声入力は特定話者で、アナウンサーによる雑音を遮断した環境での文節発話である。言語処理の解析で一つの文候補まで絞りこんだ時点での文認識率は92%である。問合せの生じた場合は不正解として数えている。

計算機条件は以下のようである。

音声認識処理は高速な演算が要求されるので、マルチプロセッサであるAlliant上で実行している。言語はCとFortanを使用している。プログラムの移植性を考えて、全般的にはCで記述している。言語処理は複雑な処理をする必要があり、文法やプログラムの開発環境を重視して、計算機はSymbolics、言語はCommonLispを使用している。統合処理は、それほど重くない処理であり、結果の表示などマンマシンインタフェースが重要であるので、ワークステーションタイプの計算機であるMicro VAXを使用し、言語は移植性を考慮しCを使用している。

### 6. おわりに

本システムでは、文節の構文情報、係り受け情報、本格的な言語処理といったように段階的に言語情報を利用し効率の良い音声言語処理を行っている。また文節の構文情報が直接的に音声認識処理に取り込まれていることに特徴がある。

また、音声認識結果の信頼性の評価を行い、音声の再入力を用いる枠組みを用意している。

今後は語彙数を順次3000語程度まで増やし、翻訳実験を進める。

#### <参考文献>

- 1] 横松、自動翻訳電話の研究動向、信学会誌Vol71, No.8解説 A. 1988
- 2] 北川端・斎藤、HMM音韻認識と予測LRパーザを用いた文節認識、信学会SP88-88、1988
- 3] 小倉・橋本・森元、言語データベース統合管理システム、情報処理学会88-NL-69-4、1988
- 4] 飯田・小暮・吉本・相沢、An Experimental Spoken Natural Dialogue Translation System Using a Lexicon-driven Grammar、Eurospeech-89、1989