

# 旅行対話のためのコーパスベース音声翻訳の評価

An Evaluation of Corpus-Based Speech Translation Systems for Travel Conversation

佐々木 裕, 菊井玄一郎, 清水 徹, 中村 哲

Yutaka Sasaki, Geníchiro Kikui, Tohru Shimizu, and Satoshi Nakamura

ATR 音声言語コミュニケーション研究所

〒 619-0288 京都府けいはんな学研都市光台 2-2-2

## 1 ATR 音声翻訳システム

本稿では, ATR 音声言語コミュニケーション研究所において, 2001 年より約 5 年間にわたって研究開発を行ってきた大規模コーパスベース多言語音声翻訳システムの性能評価結果を報告する.

ATR では, 1986 年の設立より音声翻訳の研究を行なってきており, 2001 年からは, 広範な話題に対する音声翻訳のフィジビリティを確認するために, 観光旅行の場面における旅行対話を対象に研究を進めてきた. 異なる言語対へのポータビリティ, 実環境における利用を可能にする技術の確立を目標として, 旅行対話を中心とした大規模コーパスを構築し, 大規模コーパスに基づく多言語音声翻訳システムを開発した. 対象言語ペアは, 日英, 英日, 日中, 中日である.

コーパスベース多言語音声翻訳システムの構成は, 図 1 の通りである. ユーザ端末は 8 素子マイクロフォンアレーを備えた PDA であり, ビームフォーマ, 雑音抑圧処理を施した後, 入力言語に対応した音響モデルと言語モデルを用いたデコーダで認識が行なわれる. 認識結果の信頼度に基づくリジェクション判定を行なった後, 認識結果を機械翻訳に送り, 翻訳出力を目的言語で音声合成する. 現在の実証実験用システムは, ほぼ音声入力から 1~2 秒で出力される構成となっており [1], コミュニケーションとして差し支えない速度まで高速化されている. 本稿では, ATR 音声翻訳システムによる訳文を主観評価及び自動評価した結果について報告する. 日英音声翻訳の性能を人間の TOEIC スコアに対応させた評価結果については安田ら [2] で別に報告する.

## 2 多言語コーパス

コーパスベース音声対話翻訳技術の開発と評価のために BTEC, MAD, FED と呼ばれるコーパスを収集・構築した [3]. コーパスサイズは, 日・英が約 100 万文, 日・中が約 50 万文である.

BTEC(Basic Travel Expressions Corpus) は旅行会話の専門家が旅行の様々な場面で頻繁に使われるような表現を作文して対訳を付与したコーパスである.

収集した表現の一部には読み上げによる音声データが付与されている.

MAD(Machine Translation Aided Dialogue) は, 異なる言語を話す 被験者 が実験室内で ATR の音声翻訳システムを介して行なった対話を収録したものである. 音声データを書き起こしたものに対訳を付与し, 開発セット/テストセットとして利用できるように整備した.

FED(Field Experiment Dialogue) は関西空港等において 外国人旅行者 と日本人案内者が音声翻訳を介して行なった対話を収録したものである. MAD と同様に開発セット/テストセットとして利用できるようにデータ整備を行なった.

## 3 評価結果

### 3.1 音声認識

音声認識は, 学習データ量に応じて最適な状態数を割り当てる最小記述長による隠れ状態網音響モデルと, 品詞などの単語クラスの確率を用いるクラス言語モデル, 前方文脈と後方文脈とを別々に考慮する複合 N-gram 言語モデルを用いている. これらのモデルを, 地方アクセントを考慮した日本語話者 4000 人, 英語, 中国語各約 500 人の音声コーパス, BTEC, MAD のテキストを用いて学習を行なっている. 認識語彙は約 3 万 5 千語, 発音の変形も入れると 5 万語である

評価は, 開発セットにおいて, 特定の Real Time Factor (RTF) 値に対し, 最高精度が得られるよう, 認識器のパラメータ (ビーム幅, 言語重みなど) を調整する. そのパラメータを用いて評価セットで得られた結果を最終結果としている. 詳細は, [4] を参照されたい.

音声認識ツールおよび認識器としては, ATR-SLC で開発を行なっている ATRASR を用いた [5]. また, CPU は Xeon 2.8GHz で統一して実験を行なった.

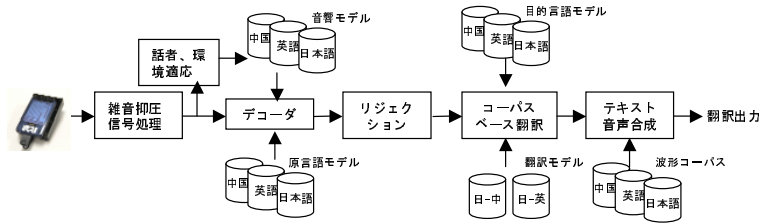


図 1: 音声翻訳システムのブロック図

### 3.2 機械翻訳

各コーパスのテストセットをマルチエンジン機械翻訳システム ATR-C3 [6] が翻訳した結果を主観評価及び自動評価した結果を示す。主観評価は 3 人の被験者により<sup>1</sup>、訳文を A(優), B(良), C(可), D(不可) の 4 段階で評価するものである。

| ランク | 基準                     |
|-----|------------------------|
| A   | Good: 良い訳              |
| B   | Fair: ほぼ問題のない訳         |
| C   | Acceptable: 大体の内容が伝わる訳 |
| D   | Nonsense: 訳文を理解できない    |
| NIL | No Output: 訳文なし        |

自動評価指標には BLEU, NIST, mWER を用いた。評価には, BTEC 510 文, MAD 502 文, FED 150 文のテストセットを用いた。各文には 16 種類の参照訳が用意されている。音声翻訳の評価結果を図 2, 3, 4 に示す。

図 2 は, 日→英翻訳における主観評価結果である。以下の条件により比較した。(1)BTEC, (2)MAD, (3)FED のテストセットそれぞれについて,

- 音声認識 (RTF=1) の翻訳 (速度重視)
- 音声認識 (RFT=5) の翻訳 (品質重視)
- テキスト翻訳 (品質重視)

の組み合わせについて評価した (グラフはこの順)。速度重視の翻訳エンジンは, 品質重視の翻訳エンジンの翻訳時間を 1 秒に制限したものである。図中の SA は音声認識の Sentence Accuracy, WA は音声認識の Word Accuracy を示す。

BTEC, MAD, FED のテキスト翻訳では C 以上の評価を合計した”ABC 率”は, それぞれ 92%, 83%, 79% であった。音声認識結果 (RTF=5) では, それぞれのコーパスについて ABC 率が約 7% 低下した。一方, 品質重視のテキスト翻訳に比べ, 速度重視型の音声翻訳では 10~20% の ABC 率の性能低下が見られた。今後, 時間あたりのシステム全体の処理性能を向上させることにより, 実証実験の場面で利用する速度重視のシステムの音声翻訳品質が改善することが明らかになった。

<sup>1</sup>性能重視システムの被検者は 1 人。

さらに, 主観評価でのテキスト翻訳の ABC 率と音声認識結果の ABC 率に注目しよう。

$$\text{推定値} = \text{テキスト翻訳の ABC 率} \times \text{単語音声認識率}$$

と定義し, 実際の音声入力の場合の測定値と比較したものが図 3 である。音声を入力とした音声翻訳の性能を音声認識率とテキスト翻訳の性能からある程度推定できることがわかる。

図 4 に自動評価結果を示す。NIST のスケールはグラフ右の値である。BLEU, NIST は大きいほど良く, mWER は値が小さいほど良い。コーパスベースアプローチにより実現された, 日英, 英日, 日中, 中日翻訳は同様な傾向を示している。FED では, 実証場面での音響的な特性 (ノイズ等) により音声認識率が低下しているが, テキスト翻訳では MAD よりも良い性能が出ている。

## 4 まとめ

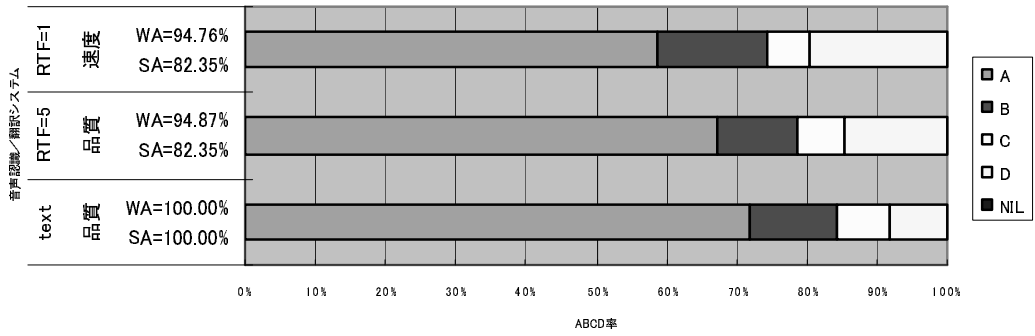
これまで当研究所で進めてきたコーパスベース音声翻訳の研究の成果について述べた。今後, さらに実用化に向けた研究開発および同時通訳に向けた基礎研究を続けていきたい。

[謝辞] 隅田主幹研究員, 今村主任研究員をはじめとする当研究所のメンバに感謝する。本研究は (独) 情報通信研究機構および総務省ネットワークヒューマンインタフェースの研究委託により実施したものである。

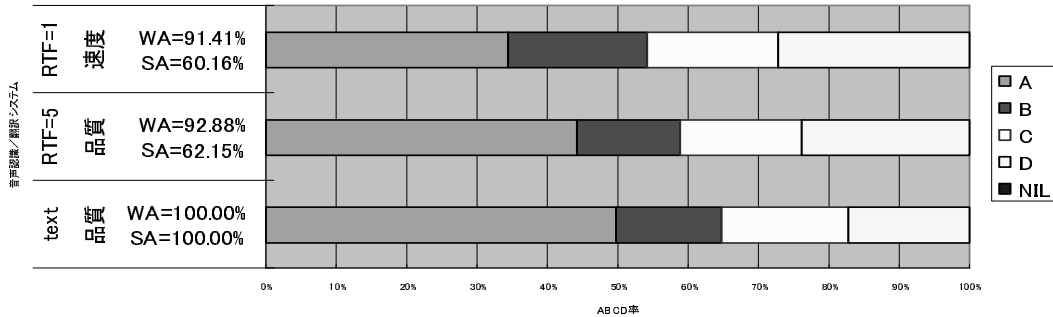
## 参考文献

- [1] 菊井 他, 信学技報 MoMuC2005-32, 2005-09.
- [2] 安田圭志, 菊井玄一郎: 人間の能力との比較による音声翻訳システム性能評価法の効率化に関する検討, 第 12 回言語処理学会年次大会, P3-3 (2006).
- [3] 菊井他, 音講論 (春), pp.55-56, 2004.
- [4] 中村 哲, 菊井玄一郎, 佐々木 裕, 清水 徹: 大規模コーパスベース音声翻訳技術と全体性能の評価, 日本音響学会 2006 年春季研究発表会, 2-1-9, pp. 87-88 (2006).
- [5] 伊藤 他: 音講論 (秋), 1-P-30, 2004.
- [6] 隅田英一郎, Paul Michael, 今村賢治, 大熊英男: 多言語音声翻訳のためのマルチ・エンジン翻訳技術, 第 12 回言語処理学会年次大会, E4-2 (2006).

(1)音声翻訳主観評価(BTEC)



(2)音声翻訳主観評価(MAD)



(3)音声翻訳主観評価(FED)

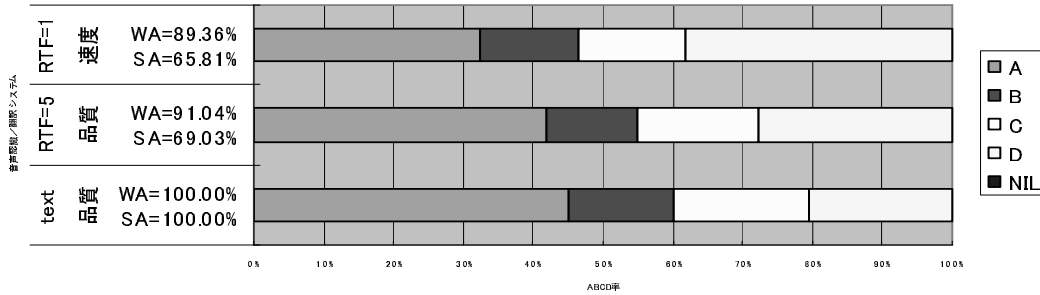


図 2: 主観評価結果 (日→英)

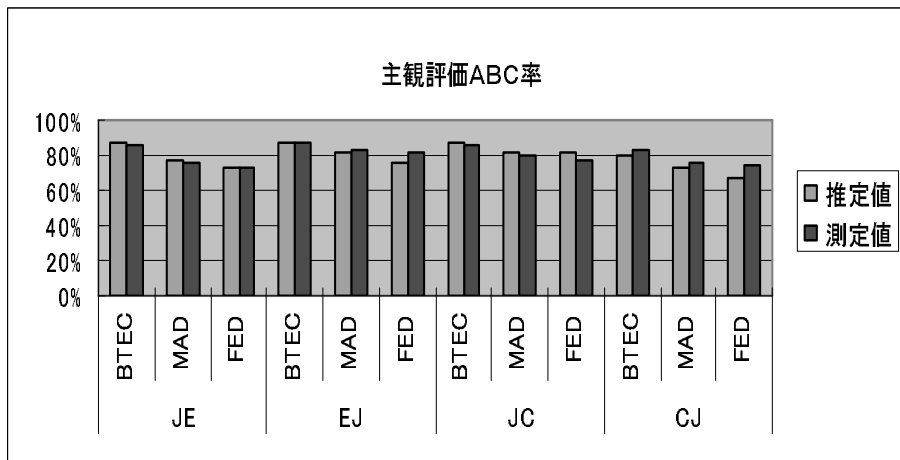


図 3: 音声翻訳の主観評価値 (ABC 率) の推定

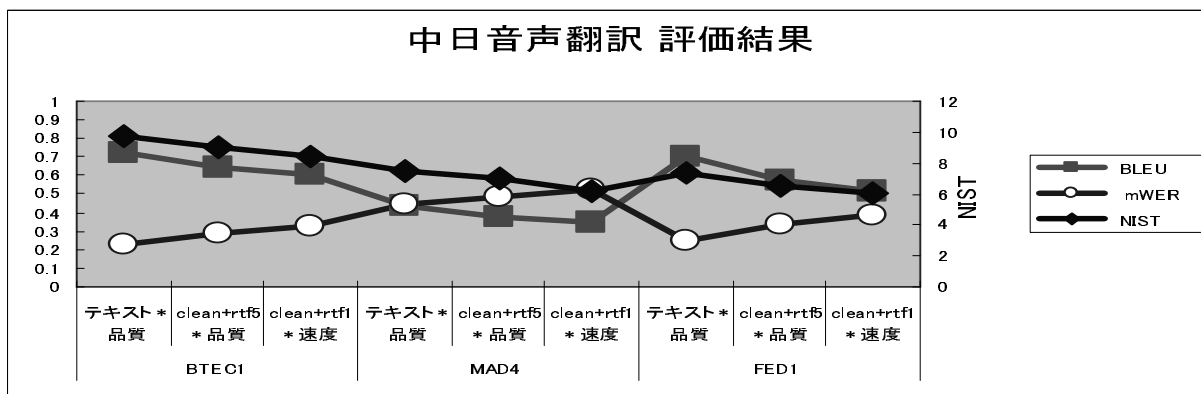
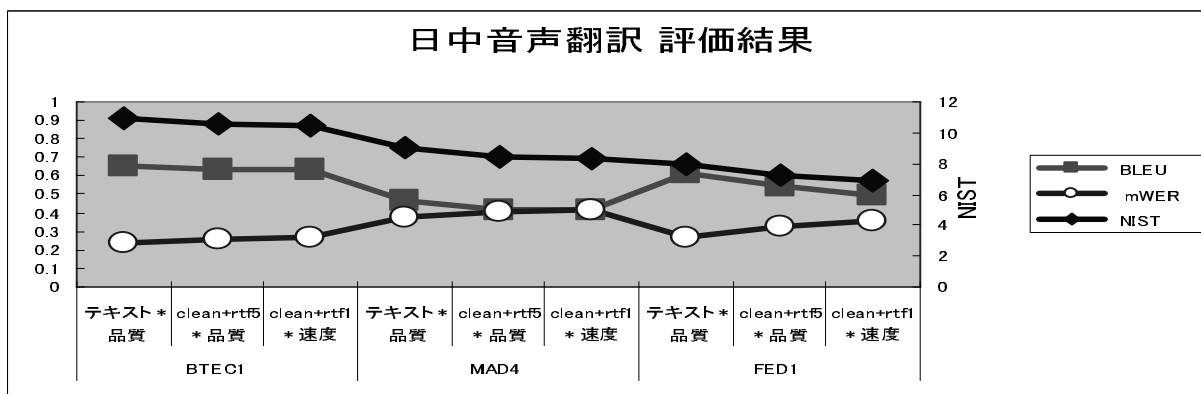
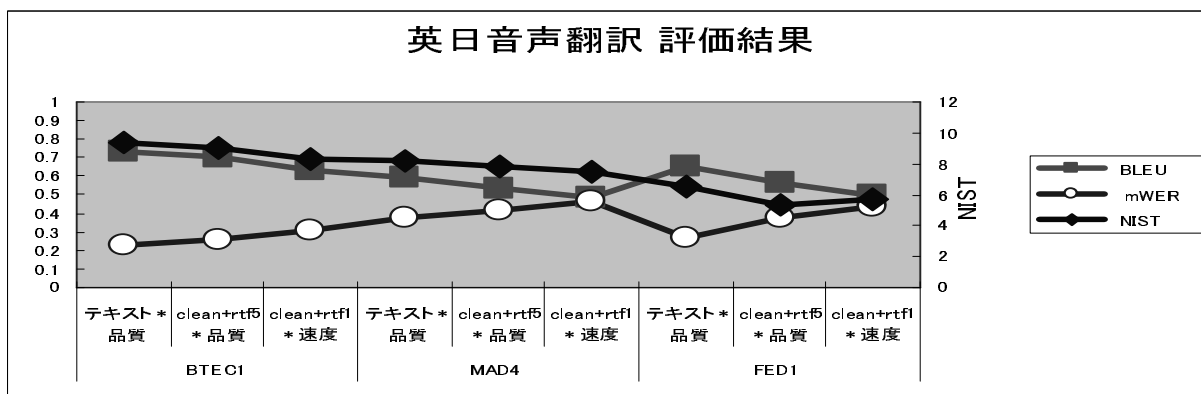
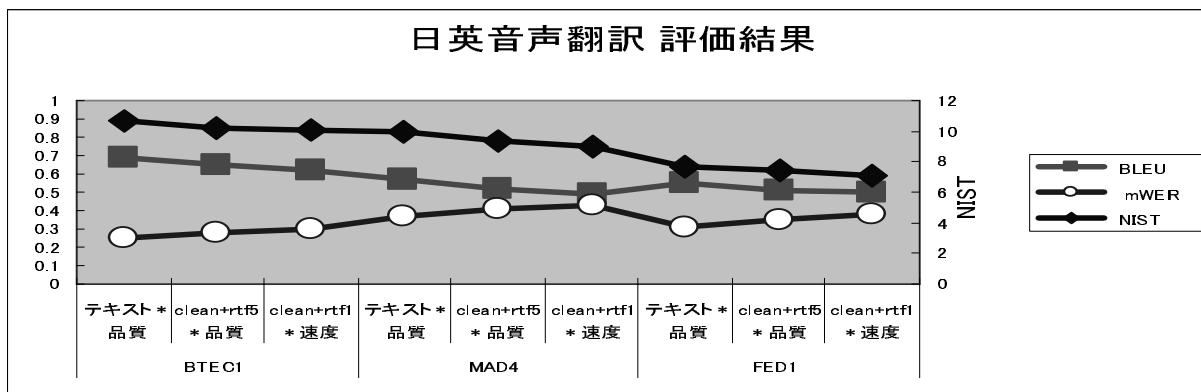


図 4: 自動評価結果 (日→英, 英→日, 日→中, 中→日)