Vol.2009-SLP-77 No.21
2009/7/18

重み付き有限状態トランスデューサーと対数線形モデルを用いた話し言葉の整形

Graham Neubig [†1]　森　信介[†1]　河原達也[†1]

自然な話し言葉には，フィラーや言いよどみ，口語的表現，助詞の脱落など，書き言葉に出現しない現象が多く含まれている．音声認識の結果から読みやすい書き起こしを作成する際に，このような現象に対応し，書き言葉に整形する必要がある．本研究では，話し言葉の整形を統計的機械翻訳の問題として扱い，この枠組みの中で様々なモデルを提案する．特に，話し言葉固有の現象を捉える複数の特徴量を組み合わせる対数線形モデルを導入する．このモデルを重み付き有限状態トランスデューサー (WFST) を用いて実装した．国会審議の書き起こしを対象とした評価実験において，ルールベースや単純な統計的モデルを用いた手法に比べて提案手法は精度を大幅に改善することができた．

Log-Linear Speaking-Style Transformation using Weighted Finite State Transducers

Graham Neubig ,[†1] Shinsuke Mori [†1]
and Tatsuya Kawahara[†1]

In spontaneous speech, there exist a number of fillers, disfluencies, colloquial expressions, and omitted words that do not appear in normal written text. In order to create easily readable transcripts from automatic speech recognition results, it is necessary to transform verbatim speech into a more appropriate written form. This paper proposes a system to correct these spoken-language phenomena based on the framework of statistical machine translation. A number of potentially useful features are proposed, and combined in a log-linear framework. The system was implemented using weighted finite state transducers (WFSTs), allowing for easy integration with WFST-based speech recognition engines. In an evaluation on a corpus from the National Diet (Congress) of Japan, the proposed system showed a large increase in accuracy over a rule-based baseline and simple statistical systems.

1. まえがき

従来の自動音声認識 (ASR) システムは音声の音響的特徴 X から，発話された内容 V を忠実に書き起こすことを目的としている．統計的手法を用いる場合，音響的特徴 X から事後確率 $P(V|X)$ が最も高くなる $\hat{V}$ を探索する．しかし，自然な発話にはフィラーや言いよどみ，口語的表現，助詞の脱落などが存在するため，忠実な音声認識結果は人間にとってかえって読みづらい．このため，プロの速記者は話し言葉に含まれる冗長性や口語的表現を修正し，読みやすい文章を作成する．音声認識を用いた書き起こしシステムも同様にこのような現象を修正し，可読性の高い文を出力することが理想的である．このような処理は人間の読者のためだけでなく，音声翻訳などの機械処理においても有益である[1)]．

話し言葉の整形はこのように自動書き起こしシステムの実現に向けて大きな課題であり，多くの研究がなされてきた．英語等の場合では，フィラーや言い直しなどの検出として定式化され，雑音のある通信路や隠れマルコフモデル (HMM)，条件付確率場 (CRF) などで整形を行うことが多い[2)–4)]．これにより冗長な表現を削除することができるが，人手による整形においては，削除のみならず脱落された単語の挿入や口語的表現の置換などの修正も行われる．これは日本語のような，話し言葉と書き言葉が著しく異なる言語では特に重要となる[5),6)]．

先行研究では，話し言葉整形を統計的機械翻訳の問題とみなし，雑音のある通信路モデルを用いて言語モデルと翻訳モデルの確率に分解している．これに対し，近年の統計的機械翻訳では対数線形モデルを採用することで，雑音のある通信路モデルで導入困難であった特徴量を扱うことが可能になっている．本研究ではこの枠組みを用いて，整形を必要とする箇所の文脈依存性や共起傾向などを考慮したモデル化を提案する．また，システムの実装法として重み付き有限状態トランスデューサー (WFST) を利用する．これにより，特徴量の開発・導入を迅速に行うことができ，WFST の音声認識システムとの統合も容易となる．

国会の会議録作成で評価実験を行い，提案手法の有効性を検証する．具体的には，人手による忠実な書き起こしと国会の会議録の “対訳コーパス” を学習・評価データとし，会議録を正解とみなしシステムの出力の単語誤り率 (WER) を評価基準として評価を行う．

†1 京都大学 情報学研究科
Graduate School of Informatics, Kyoto University

Vol.2009-SLP-77 No.21
2009/7/18

発話	いろんな	えー	ところ	で	注文		つける	と	です　ね	…
会議録	いろいろ　な		ところ	で	注文	を	つける	と		…
	置換	フィラー				挿入			フィラー以外	

図 1　話し言葉整形の一例
Fig. 1　An example of transformed speech

2. 話し言葉の整形

会議録や講演録を作成する上で編集者が行う修正として以下のようなものがある[6]．

(1)　フィラー等の削除: 「あのー」や「えーと」などのフィラー単語を削除する．「ですね」のように，フィラーとしても機能語としても利用される語についてはフィラーか否か判断する必要がある．

(2)　書き言葉表現への変換: 「いろんな」などの口語的表現を「いろいろ な」のような文書にふさわしい書き言葉表現に変換する．

(3)　助詞の挿入: 「注文つける」のように助詞が脱落された文に「注文 を つける」のように適切な助詞を挿入する．

(4)　その他の修正: 倒置部や言い淀み，独り言，単語の誤用，誤った情報の修正を行う．

整形前と整形後の文の例を**図 1** に示す．本研究では統計的手法を用いて，主に (1)〜(3) を対象とする．

3. 整形のモデル化

話し言葉の整形を行うために，発話の忠実な書き起こし V と会議録 W を異なる言語とみなし，V から W への統計的機械翻訳を行う．事後確率 $P(W|V)$ を以下のような手法でモデル化し，与えられた V に対して $P(W|V)$ が最も高くなる $\hat{W}$ を探索する．

3.1 雑音のある通信路モデル

統計的機械翻訳は原言語と目的言語が文毎に対応付けされている対訳コーパスを必要とする．対訳コーパスの量は単言語で入手可能な言語資源の量をはるかに下回るため，ベイズ則を用いて $P(W|V)$ を言語モデル確率 $P(W)$ と翻訳モデル確率 $P(V|W)$ に分解する．

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(V|W)P(W) \tag{1}$$

翻訳モデル確率 $P(V|W)$ の学習には対訳コーパスが必要となるが，言語モデル確率 $P(W)$ は大量の単言語データから学習可能である．

翻訳モデル確率を推定する際に，単語の翻訳確率が文脈に依存しないと仮定し，文の翻訳モデル確率を各単語の翻訳確率の積で近似する．

$$P(V|W) \approx \prod_i P(v_i|w_i) \tag{2}$$

$P(v_i|w_i)$ を算出するために，原言語文 $V = v_1, \ldots, v_k$ と目的言語文 $W = w_1, \ldots, w_k$ をアラインメントし，アラインメントされた単語対 $\gamma_i = \langle v_i, w_i \rangle$ の列 $\Gamma = \gamma_i, \ldots, \gamma_k$ を作成する．γ_i と w_i の出現回数 $c(\gamma_i)$ と $c(w_i)$ を数え，以下の式で $P(v_i|w_i)$ を求める．

$$P(v_i|w_i) = \frac{c(\gamma_i)}{c(w_i)} \tag{3}$$

削除や挿入の確率を計算するために，ヌル文字列 ϵ を用いて，削除・挿入された単語に対して，それぞれ $P(v|\epsilon)$ と $P(\epsilon|w)$ を推定する．「いろんな」→「いろいろ な」のような一対多関係を扱うために「いろいろ な」を 1 つの単語として語彙に登録する．

統計的機械翻訳の手法で話し言葉の整形を行う先行研究のほとんどはこの雑音のある通信路モデルを用いている[2),3),5),6)]．

3.2 対数線形モデル

話し言葉の整形では，言語モデル確率と翻訳モデル確率以外にも有用な特徴がある（第 4 節参照）．このような特徴量を導入する方法として対数線形モデルが効果的である[7)]．簡単な対数線形モデルは式 (1) の対数を取り，それぞれの項に重みをつけることによって得られる．

$$\hat{W} = \underset{W}{\operatorname{argmax}} \left[\lambda_1 \log P(V|W) + \lambda_2 \log P(W)\right] \tag{4}$$

λ_1 と λ_2 の値が両方 1 である場合，式 (1) と式 (4) で得られる $\hat{W}$ は同一である．したがって，最適な重みを選択すれば，対数線形モデルにより単純な雑音のある通信路モデルより高い精度が期待できる．重みの学習については第 5.4 節で述べる．

また，対数線形モデルにおいて，多様な特徴量を組み合わせることも可能である．単純な雑音のある通信路モデルと違い，特徴量は独立ではなくてもよいため，新しい特徴量の導入・検証が容易となる．

3.3 翻訳モデルの分解

雑音のある通信路における単語翻訳確率 $P(v_i|w_i)$ は式 (3) のように γ_i と w_i の出現回数から算出される．これを利用して，以下のように翻訳確率をその 2 つの頻度の対数を分解し

た形で表現できる．

$$\begin{aligned}\log P(V|W) &= \log \prod_i P(v_i|w_i) \\ &= \log \prod_i c(\gamma_i)/c(w_i) \\ &= \log \prod_i c(\gamma_i) - \log \prod_i c(w_i)\end{aligned}$$

この 2 つの対数頻度を対数線形モデルにおいて個別の要素として扱えば，重み学習の段階で両方に異なる重みがつけられる．前者の重みが後者の重みより大きい場合，学習コーパスに頻出する変換は頻度の低い変換より重視され，適合率の向上が期待できる．以下ではこのモデルを “分解モデル” と呼ぶ．

3.4 文脈依存の翻訳モデル

雑音のある通信路では，$P(W|V)$ を文脈に依存する言語モデル $P(W)$ と文脈に依存しない翻訳モデル $P(V|W)$ に分解する．$P(W)$ の学習には大規模な単言語データが利用できるという利点もあるが，$P(V|W)$ の文脈非依存性は必ずしも成り立たない．特に，「ですね」や「と」のような，変換するか否かが文脈に依存する現象については，翻訳モデルに直接文脈を取り入れた方が有利であると考えられる．

文脈に依存する翻訳モデルを作成するために，雑音のある通信路モデルでモデルの分解を行わずに，対訳コーパスのみを用いて同時確率 $P(V,W)$ を直接モデル化する方法がある．

$$\begin{aligned}\hat{W} &= \underset{W}{\operatorname{argmax}} P(V|W)P(W) \\ &= \underset{W}{\operatorname{argmax}} P(V,W)\end{aligned}$$

同時確率をモデル化する具体的な方法はいくつか提案されているが，本研究では GIATI[8] と呼ばれる方法を採用する．アラインメントされた V と W の組を表す Γ を用いて，平滑化された n-gram モデルを構築する．

$$P(V,W) = P(\Gamma) \approx \prod_i^k P(\gamma_i|\gamma_{i-n+1},\ldots,\gamma_{i-1})$$

これにより翻訳モデルに文脈を利用することはできるが，同時確率を直接モデル化するため，言語モデル $P(W)$ との併用は容易ではなく，単言語のデータを利用することは困難である．

4. 整形のための特徴量

本節では，話し言葉整形の特徴を考慮した特徴量について述べる．

4.1 フィラー辞書

話し言葉整形の最も単純な形として，人手で作成されたフィラーの辞書に含まれている単語をすべて削除する方法がある．文脈に依存しないフィラーしか削除できないため修正の再現率は良くないが，実装が容易であり，よく用いられる方法である．フィラー辞書の情報を対数線形モデルで使用するために，フィラーが削除される度に定数のボーナスを加える特徴量を導入する．フィラー辞書の特徴量の重みのみを 1 にすることで対数線形モデルにおいてルールベースシステムと同じ動作が再現できる．このため，重みを適切に選択すれば，対数線形モデルの精度はルールベースシステムの精度以上になることが期待される．

4.2 変換グループ

話し言葉の中で，整形が必要となる箇所は連続して出現することが多い[⋆1]．これは主に，話し手が思考している間，フィラーや言い直しなどが連続的に発声されるからである．

この現象を捉える特徴量として，変換グループのペナルティーを提案する．連続して変換されるグループ毎に，グループのサイズに係わらず，定数のペナルティーを加算する．これで「えー です ね えー」のように 2 変換グループの「えー です ね えー」としても，1 変換グループの「えー です ね えー」としても変換され得るものがあれば，1 つの変換グループの後者の方が優先的に選ばれ，変換の共起性が捉えられる．

4.3 変換タイプ

話し言葉の整形で必要となる修正は，削除・挿入・置換に分類できる．第 6 節の評価実験で明らかとなるように，削除の精度は挿入や置換の精度より高くなる傾向がある．削除・挿入・置換の 3 タイプに個別のペナルティー（ボーナス）を与えることで，特定が容易な変換はより多く行われ，特定困難な変換については抑制される．

4.4 単 語 数

音声認識や機械翻訳では，特に長いまたは短い候補が探索に障害を及ぼすことを防ぐために，仮説に含まれる単語毎に定数のペナルティーを加算することが多い．ここでも単語数に応じてペナルティーを導入する．

⋆1 評価用データの中で変換すべき単語は全体の 16.2%であるため，変換すべき単語がランダムに起こると仮定すれば変換グループの長さの期待値は $1/(1-0.162) = 1.19$ 単語である．これに対して，変換グループの実際の平均長は 1.76 単語であるため，ランダムに起こるわけではないといえる．

Vol.2009-SLP-77 No.21
2009/7/18

5. モデルの構築と探索

5.1 重み付き有限状態トランスデューサー

以上のモデルを重み付き有限状態トランスデューサー (WFST) で実装した．WFST は有限オートマトンの拡張であり，各状態遷移は入力・出力・重みを有する．入力列に従って状態遷移を繰り返し，その結果，出力と重みが得られる．詳細は文献 9) を参照されたい．

状態遷移の重みが確率（通常は負の対数確率）に相当する場合，$P(V|W)$ や $P(W)$ などの確率的モデルを WFST で表現することができる．複数の WFST を効率的に合成するアルゴリズムや，1 つの WFST を決定化・最小化するアルゴリズムなどが知られているため，個別にコンポーネントを作成しても容易に 1 つのシステムに組み合わせることができる．これで特徴量の効率的な開発が可能となり，また WFST により実装された音声認識システムと組み合わせれば音声から直接会議録や講演録を作成するシステムも構成できる．

5.2 モデル学習

本システムを構築するために，第 4 節で述べた各特徴量を個別の WFST で表現する．このため，ほとんどの WFST（ペナルティーや変換候補作成等）を構築するのはさほど困難ではない．言語モデルや GIATI による翻訳モデルを学習するために通常の n-gram ツールキットを利用することができる．後述する評価実験では，WFST が直接出力できる言語モデルツールキット Kylm [⋆1]を利用した．

後述する評価実験では，言語モデルの学習には Kneser-Ney 平滑化[10] を用いた．言語モデル確率 $P(W)$ は 3-gram で推定し，GIATI による翻訳モデルは n を変化させて，精度を比較した．

5.3 WFST の合成と探索

各特徴量を WFST で表現してから，それらを合成することで 1 つのモデルが容易に構築できる．また，最小化・決定化などで，必要な記憶量を削減でき，探索が効率的となる．本研究ではこれらのアルゴリズムを実装したライブラリ OpenFst[11] を利用した．

合成されたモデル上で全探索を行うことは計算量的に現実的でないため，ビームサーチを用いる．ビームサーチを行うために，OpenFst をベースとした WFST デコーダを開発した[⋆2]．探索は 2 パスで行い，まずビームサーチで探索空間を枝刈りしてから全探索で枝刈りされた空間の中の最適な解を探し出す．

⋆1 第一著者作のオープンソースツールキット．http://www.phontron.com/kylm にて入手可能．
⋆2 今後オープンソースで公開する予定である．

表 1 評価用データの詳細
Table 1 Test set details

文数		7,181	
単語数	発話の書き起こし	341,398	
	会議録	299,214	
削除	フィラー	27,510	(49.1%)
	その他	18,789	(33.6%)
置換		5,558	(9.9%)
挿入		4,115	(7.4%)
単語誤り率			18.71%

表 2 各翻訳モデルの精度
Table 2 Accuracy of various translation models

モデル		WER	+全特徴量
整形なし		18.71%	-
ルールベース		10.86%	-
雑音のある通信路		6.91%	-
+重み学習		6.40%	5.74%
+モデル分解		5.85%	**5.37%**
GIATI	1-gram	9.95%	7.36%
	2-gram	5.56%	5.29%
	3-gram	5.50%	**5.04%**
	4-gram	5.62%	5.13%

5.4 重みの学習

対数線形モデルの最適な重みを学習する方法として，誤り率最小化学習法 (MERT) がある[12]．学習コーパスから少量のデータを重みの学習のために取っておき，そのヘルドアウトデータに対する誤り率が最小となる重みを学習する．誤り率の定義は任意であるが，本研究では評価基準としても利用する単語誤り率 (WER) を採用する．評価実験ではオープンソースの機械翻訳ツールキット Moses[13] に含まれている誤り率最小化学習のツールを利用した．

6. 評 価 実 験

6.1 実験データ

学習・評価用データとして衆議院の会議録を用いた．言語モデルの学習に 1999 年 1 月〜2007 年 8 月の会議録から 364 万文，翻訳モデルの学習に 2003 年 2 月〜2006 年 10 月の 5.62 万文の対訳コーパスを使用した．語彙は国会の音声認識のために作成された 49,270 語の単語辞書で設定し，重みは 974 文のヘルドアウトデータで学習した．

2007 年 10 月に行われた会議 7 回の忠実な書き起こしと会議録を評価の対象とした．評価用データの詳細は表 1 の通りである．会議録を正解とみなして，これと比較して得られる単語誤り率 (WER) を計算する．削除の欄で「フィラー」とされるものは第 4.1 節で言及したフィラー辞書を用いてルールベースで削除できるものである．ここでは 23 語のフィラーを登録した．フィラーの削除はすべての変換の 49.1%を占め，最も多い変換ではあるが，必要な変換の半分以上はフィラーの削除ではないため，統計的手法は明らかに必要である．

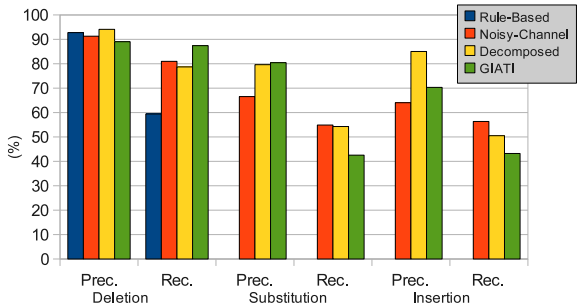

図 2 各変換タイプの適合率と再現率
Fig. 2 The precision and recall of each transformation type

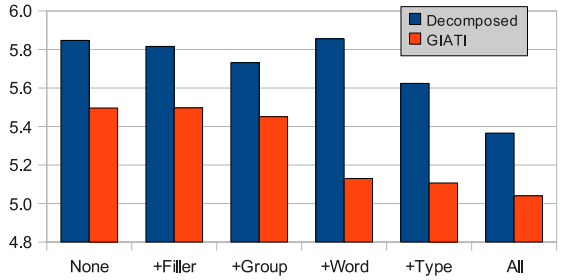

図 3 各特徴量を用いた際の単語誤り率
Fig. 3 The WER when each feature is used

6.2 翻訳モデル

まず，第 3 節で説明した翻訳モデルの比較を行った．ベースラインとして，ルールベースのフィラー削除と重み学習なしの雑音のある通信路モデルを用いた．GIATI に基づくモデルの n-gram 長を 1〜4 の間で変化させた．

それぞれの翻訳モデルの精度は表 2 の通りである．ルールベースモデルではフィラーの削除のみができ，整形なしより単語誤り率を 7.85%改善できたが，その他の変換を全く行わず，モデルの中で最も単語誤り率が大きかった．単純な雑音のある通信路モデルを用いて統計的に処理を行うことで，大幅に単語誤り率が削減できている．$P(W)$ と $P(V|W)$ の個別の重みを学習することで精度を向上させることができ，翻訳モデルを分解することでさらに有意な改善が見られた．

GIATI については，文脈を考慮しない 1-gram モデル以外は雑音のある通信路モデルと分解モデルより高い精度となった．GIATI の中で最も誤りが少なかったのは 3-gram モデルであったため，以下の分析では 3-gram モデルを GIATI として言及する．

各変換タイプにおける適合率と再現率を**図 2** に示す．まず，分解モデルと通常の雑音のある通信路モデルを比較すると，モデルを分解することですべての変換タイプにおいて適合率が大幅に改善され，再現率がわずかに下がる傾向が見られた．第 3.3 節で述べたように，$\prod_i c(v_i, w_i)$ の重みを上げることで信頼度の高いパターンが優先的に変換されるためであると考えられる[⋆1]．また，分解モデルと GIATI を比較すると，GIATI は削除の再現率において分解モデルを大幅に上回ったものの，それ以外ではほぼ同等，または低い精度となっていた．削除パターンが最も高い頻度であるので，文脈情報を効果的に学習することができ，全体の WER に対する貢献も大きい．その半面，置換や挿入のデータは比較的少なく，文脈の学習は困難であると考えられる．

6.3 特徴量の効果

次に，対数線形モデルにおける各特徴量の効果を検証するために，特徴量を 1 つずつ導入し，その際の単語誤り率を比較した．その結果を**図 3** に示す．特徴量の効果はそれぞれ以下の通りである．

- **フィラー辞書：**分解モデルでも GIATI でも大きな効果がなかった．これは辞書に入っているフィラーは特徴量を個別に与えなくてもきちんと処理されていたからである．
- **変換グループ：**分解モデルで大きな改善が見られたが，GIATI における改善は比較的小さかった．分解モデルでは隣の単語が変換されたかどうかは有益な情報となるが，GIATI では文脈情報が翻訳モデルに含まれているためあまり有益でないと考えられる．
- **単語数：**GIATI では大きな改善が見られたが，分解モデルでは全く効果がなかった．図 2 で明らかなように，GIATI は分解モデルに比べて削除の再現率が高かったが，適合率は低かった．単語数でペナルティーをかけることで適合率をあげ，精度が大幅に改善された．これに対して，分解モデルは既に適合率と再現率のバランスが取れていたため，単語数のペナルティーは効果がなかった．
- **変換タイプ：**分解モデルでも GIATI でも各タイプの適合率と再現率のバランスを調整し，適切な値を選択することで，すべての特徴量の中で最大の改善が得られた．

⋆1 $\prod_i c(v_i, w_i)$ と $\prod_i c(w_i)$ と $P(W)$ の重みはそれぞれ 1.64，-1.27，1.00 であった．

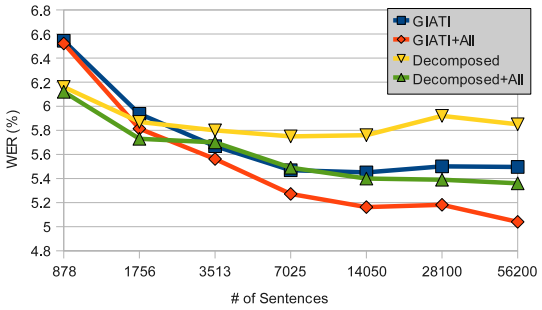

図 4 学習に用いる対訳コーパスの文数と単語誤り率の比較
Fig. 4 A comparison of word error rates for different sizes of the parallel corpus

また，すべての特徴量を組み合わせることで，精度はさらに改善し，GIATI のモデルでは 5.04%，分解モデルでは 5.37%の単語誤り率が得られた．

6.4 コーパスサイズの影響

忠実な書き起こしの準備には大きなコストがかかり，実用的には少ないデータ数で高い精度が得られることが求められる．このため，学習に用いる対訳コーパスのサイズを変化させて，分解モデルと GIATI の精度を比較した．図 4 にこの比較の結果を示す．

分解モデルでも GIATI でも，コーパスのサイズが小さくなるほど精度が下がるが，GIATI の方が大きな低下を見せ，分解モデルを下回ることがわかる．これは，GIATI は対訳コーパスのみを利用しているのに対し，分解モデルは会議録のみのコーパスも利用しているからである．また，コーパスのサイズが小さくなるほど全特徴量を用いる効果が小さくなる．最も小さいコーパスでは全特徴量を用いても用いなくてもほぼ同等の単語誤り率が得られた．

7. む す び

本稿では，話し言葉を整形する手法を提案し，様々な翻訳モデルの構造と特徴量を導入し，その効果を比較した．GIATI によるモデルは十分な学習量がある場合には文脈を捉え，最も高い精度が得られた．雑音のある通信路モデルの拡張である分解モデルも単純なモデルより高い精度が得られ，少ないデータで比較的安定して学習できた．変換タイプや変換グループ，単語数などの特徴量は有用であり，フィラーの辞書は大きな効果がなかった．すべての特徴量の組み合わせによる WER は 5.04%であり，ルールベース手法の 10.86%と単純な雑音のある通信路モデルを用いる手法の 5.71%から大幅に改善した．

参 考 文 献

1) Rao, S., Lane, I. and Schultz, T.: Improving Spoken Language Translation by Automatic Disfluency Removal: Evidence from Conversational Speech Transcripts, *Machine Translation Summit XI*, Copenhagen, Denmark, pp.177–180 (2007).
2) Honal, M. and Schultz, T.: Correction of Disfluencies in Spontaneous Speech using a Noisy-Channel Approach, *Proc. EuroSpeech2003*, Geneva, Switzerland, pp. 2781–2784 (2003).
3) Johnson, M. and Charniak, E.: A TAG-based noisy channel model of speech repairs, *Proc. ACL04*, Association for Computational Linguistics, p.33 (2004).
4) Liu, Y., Shriberg, E., Stolcke, A., Hillard, D., Ostendorf, M. and Harper, M.: Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.14, No.5, pp.1526–1540 (2006).
5) Hori, T., Willett, D. and Minami, Y.: Language model adaptation using WFST-based speaking-style translation, *Proc. ICASSP2003*, pp.228–231 (2003).
6) 下岡和也，南條浩輝，河原達也：講演の書き起こしに対する統計的手法を用いた文体の整形，自然言語処理， Vol.11, No.2, pp.67–83 (2004).
7) Och, F.J. and Ney, H.: Discriminative training and maximum entropy models for statistical machine translation, *Proc. ACL02*, Morristown, NJ, USA, Association for Computational Linguistics, pp.295–302 (2002).
8) Casacuberta, F. and Vidal, E.: Machine Translation with Inferred Stochastic Finite-State Transducers, *Computational Linguistics*, Vol. 30, No. 2, pp. 205–225 (2004).
9) Mohri, M.: Finite-state transducers in language and speech processing, *Computational Linguistics*, Vol.23, No.2, pp.269–311 (1997).
10) Kneser, R. and Ney, H.: Improved backing-off for M-gram language modeling, *Proc. ICASSP95*, Vol.1, pp.181–184 (1995).
11) Allauzen, C., Riley, M., Schalkwyk, J., Skut, W. and Mohri, M.: OpenFst: a general and efficient weighted finite-state transducer library, *Proc. CIAA '07*, pp.11–23 (2007).
12) Och, F.J.: Minimum Error Rate Training in Statistical Machine Translation, *Proc. ACL03*, pp.160–167 (2003).
13) Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. and Herbst, E.: Moses: Open Source Toolkit for Statistical Machine Translation, *Proc. ACL07*, Prague, Czech Republic, Association for Computational Linguistics, pp. 177–180 (2007).