音声言語情報処理 45−12
（2003．2．8）

コーパスベース音声翻訳研究のための対話データ収集

竹澤寿幸[†]　菊井玄一郎[†]
鈴木弥生[†]　西野敦士[†]

ホテル場面のみならず，買物や食事などいろいろな旅行に関する場面で音声翻訳システムが使えるようにするために，大規模な日英対訳コーパスを作成している．その一環として，実際に翻訳システムを介して日本語話者と英語話者が課題遂行対話を行う実験を実施した．そのような状況でのコミュニケーション行動，および，その際に好まれる言語表現に関するデータを収集することが目的である．現状の音声翻訳システムをそのまま使うと制約が大きいので，音声認識器の代わりに人間(タイピスト)が発話内容を書き起こして翻訳システムに入力する形態を試みた．実験の概要を報告し，対話データが効率の良いコミュニケーションと豊かなコミュニケーションに分類できることを示す．

Collecting Machine-Aided Dialogues for Corpus-Based Speech Translation

Toshiyuki Takezawa,[†] Genichiro Kikui,[†] Yayoi Suzuki[†]
and Atsushi Nishino[†]

A huge bilingual corpus of English and Japanese is being built at ATR Spoken Language Translation Research Laboratories in order to enhance speech translation technology, so that people can use a portable translation system for traveling abroad, dining and shopping, as well as hotel situations. As a part of these corpus construction activities, we have been collecting dialogue data using an experimental translation system between English and Japanese. The purpose of this data collection is to study the communication behaviors and linguistic expressions preferred in front of such systems. Since there are a lot of limitations in state-of-the-art speech translation systems, we use human typists to transcribe the users' utterances and input them into a language translation system between English and Japanese instead of using speech recognition systems. In this paper, we present an overview and also show that the dialogue data is classified into two groups: efficient communications and rich communications.

1. まえがき

外国人旅行者とホテルのフロント係の会話のように限定された場面で実証されたコーパスベース音声対話翻訳技術[1]を広い範囲の旅行会話が扱えるように拡張したり，あるいは移植ないし適応技術の研究を進めるためには研究用のコーパスが必要である．今後の音声翻訳研究用コーパスを設計する上で検討しなければいけない重要な点が三つある．一つめは音声の多様性，具体的には発音，発話様式，話者等の多様性である．二つめは場面の多様性，具体的には話者の置かれた状況(空港，ホテル，食事，買物，旅先でのトラブル等)や，話者の役割(街角の見知らぬ人，店員等)の多様性である．三つめは表現，具体的には表層的な言い回しの多様性である．

先行研究[1]によれば，機械を意識した発話は同一言語話者同士の対話音声と朗読音声の中間的な特徴を有し，しかもユーザが慣れるにしたがって朗読音声の特徴に近づく傾向が見られる．もし音声対話翻訳の扱う対象が機械を意識した発話様式で良いとすれば，音声データの多様性の収集はバイリンガルデータの収集と分離して取り扱っても良いことになる．そうすると，バイリンガルデータ収集で検討すべき点は場面の多様性と表現の多様性の二つになる．

場面の多様性を網羅するバイリンガルデータ収集として，旅行会話基本表現集[2]を構築した．また，表現の多様性を網羅するデータ収集として，旅行会話基本表現集の基本表現に対する言い換え表現の収集を試みている[3]．

しかしながら，これら二つの試みは音声データの収集と完全に分離することで量を増やすことに重点を置いた

† (株)国際電気通信基礎技術研究所 音声言語コミュニケーション研究所
ATR Spoken Language Translation Research Laboratories

ものである．したがって，対象こそ会話調の表現であっても音声起源でないため，その特性が音声翻訳システムが本来扱うべき対象と若干異なる可能性がある．

そこで，音声翻訳研究用コーパス収集の一環として，実際に翻訳システムを介して日本語話者と英語話者が課題遂行対話を行う実験を実施した．そのような状況でのコミュニケーション行動，および，その際に好まれる言語表現に関するデータを収集することが目的である．ただし，現状の音声翻訳システムをそのまま使うと制約が大きいので，音声認識器の代わりに人間 (タイピスト) が発話内容を書き起こして翻訳システムに入力する形態を試みた．本稿では，実験の概要を報告し，対話データが効率の良いコミュニケーションと豊かなコミュニケーションに分類できることを示す．

2. では機械介在対話方式によるデータ収集の必要性を述べる． **3.** では実験システム構成を述べる． **4.** では対話実験について述べる． **5.** では議論を行う．最後に **6.** で全体をまとめる．

2. 機械介在対話データ収集の必要性

音声翻訳研究のために ATR で構築したバイリンガルコーパスは二つある．一つは旅行会話基本表現集 (BTEC: Basic Travel Expression Corpus)[2] である．もう一つはバイリンガル旅行会話コーパス (SLDB: Spoken Language Data Base)[4] である． BTEC と SLDB は相補的になっている． BTEC は音声起源でない旅行会話基本表現の日英対訳を集めたものである．海外旅行に出かける人向けに出版されている会話例文集 (いわゆるフレーズブック) に現れるような広い範囲の話題を網羅し，その規模は 20 万文を越えている．一方，SLDB は通訳者を介した日英バイリンガル会話を書き起こしたものである．ただし，ホテル場面に限定されており，その規模は 2 万文を越える程度である．

その内容を分析した報告[5] によれば，実世界で利用する場合に音声翻訳システムが扱わなければならないであろう内容と， ATR で構築したバイリンガルコーパスおよびBTEC から選んだ基本表現に基づく言い換え表現には次のような違いがあることが示唆されている．

- BTEC: 音声起源でなく，しかも言い回しが編集 (整形) されているため，日常的な会話表現と異なる場合がある．
- SLDB: 通訳者が介在しているため，認識誤りや翻訳誤りが含まれていない．しかしながら，現状ないし近未来の音声翻訳システムでは認識誤りや翻訳誤りが避けられない．
- 言い換え表現収集: できる限り数多く言い換え表現を作成させると，基本表現に基づく丁寧な言い回しばかりが多くなる傾向がある☆．

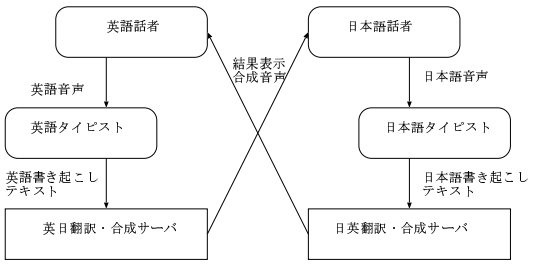

図 1 実験システム構成
Fig. 1 Experimental system configuration

実際に翻訳システムを介して日本語話者と英語話者が課題遂行対話を行うことにより得られたデータ (MAD: Machine-Aided Dialogues) を分析すれば， ATR で構築したコーパスとの定量的な特性の差を把握することができる．そのために，音声翻訳研究用コーパス収集の一環として，機械介在対話形式によるデータ収集が必要となる．また，もちろん，このような対話音声データを音声翻訳技術およびその要素技術の評価実験に利用することもできる．

3. 実験システム構成

実験システム構成を図 1に示す．英語話者の音声を英語タイピストが書き起こしテキストに変換し，英日翻訳・合成サーバに入力する．翻訳結果と合成音声が日本語話者に送られる．日本語話者の音声を日本語タイピストが日本語書き起こしテキストに変換し，日英翻訳・合成サーバに入力する．翻訳結果と合成音声が英語話者に送られる．これを繰り返すことで翻訳システムを介した対話が行われる．なお，音声波形，書き起こしテキスト，翻訳結果は日本語，英語ともにデータとしてすべてファイルに保存される．

日英翻訳には音声翻訳システム ATR-MATRIX[1] の翻訳部 TDMT (Transfer Driven Machine Translation)[6] を拡張したものに DP マッチングを用いた用例に基づく機械翻訳 D3 (DP-matching Driven transDucer)[7] を組合わせたものを使用した☆☆．英日翻訳には TDMT を拡張したものを使用した．日本語音声合成には CHATR[8] を使用した．英語音声合成には CHATR

☆ たとえば，文末表現の「ください」「くださいますか」「くださいますでしょうか」「くださいませんか」「くださいませんでしょうか」「くださらないですか」「くださらないでしょうか」「くださるでしょうか」等．

☆☆ D3 で利用する用例距離計算の値が 0.2 より小さければ D3 の結果を採用し，それ以外は TDMT を拡張したものの結果を採用する．

(第 1 回対話実験), AT&T Labs' Natural Voices^{TM} (第 2 回対話実験) を使用した.

4. 対話実験

4.1 概　　要

対話実験は 2 回のシリーズに分けて実施した. その概要は次の通りである.

- 第 1 回対話実験 (MAD1)
  - ねらい: 実現可能性の検証
  - 内容: 短い対話 (例: タクシー乗り場を尋ねる)
  - 実施時期: 2002 年 5 月から 6 月の 12 日間
- 第 2 回対話実験 (MAD2)
  - ねらい: 課題達成と発話数の関係調査
  - 内容: 少し複雑な対話 (例: ツアーを予約する)
  - 実施時期: 2002 年 8 月から 9 月の 11 日間

4.2 ユーザへのインストラクション

実験参加者であるユーザに対しては, 実験の目的を説明した後, 次の点に注意するよう指示した.

- 大きめの声で明瞭に話す.
- 1 回の発話は 10 秒以内とする☆.
- 時々誤りが発生するが, 確認や再発話をすることにより対話を続ける.
- 時々処理に時間がかかることがあるが, その場合は少し待つ.

実際の機器操作としては, ボタンを押してからユーザに話させ, 実験システムではボタンを押した後の 10 秒間のみ録音とタイピストへの転送を行った.

4.3 実験の実施

まず 1 ターンを 1 分以内で実現することを目標とした. 内訳は日本語話者発話 10 秒, 日本語タイピスト作業とシステム処理時間あわせて 10 秒, 英語合成音声出力 10 秒, 英語話者発話 10 秒, 英語タイピスト作業とシステム処理時間あわせて 10 秒, 日本語合成音声出力 10 秒である. そのような能力のあるタイピストを日本語, 英語ともにオーディションにより選び, 訓練することで, 目標は達成できた. タイピストにはなるべく忠実に発話を書き起こすよう指示した. 実際には 1 ターンが 30 〜 40 秒でなされることもあり, 時々処理に時間がかかることを除けば, おおむね人間の通訳者が介在する場合と比べて速度的には大きな差はないと言える☆☆.

第 1 回対話実験の際は, 実際の状況に近いシステム構成が望ましいと考え, 携帯電話 2 台 (日英話者それぞれ 1 台ずつ) と PDA 1 台 (共通) をユーザインタフェースに採用した. しかしながら, 現状の合成音声を携帯電話から出力する条件で得られる品質では, 出力内容確認を常に PDA の画面表示に頼ることになってしまい, 1 台の PDA を共通に使う構成では必ずしもユーザの使い勝手が良くなかった. また, 携帯電話品質の音声よりも接話型マイクの音声が現在の音声認識研究には適していることもわかった. そこで, 第 2 回対話実験シリーズでは, それぞれの話者に一つずつ小型ノート PC とヘッドホン付き接話型マイクを与える構成に変更した. このようにして収集した量は次の通りである.

- 第 1 回対話実験 (MAD1)
  - 日本語話者: 1 日あたり 2 名, 延べ 24 名 (異なり 19 名)
  - 英語話者: 1 日あたり 1 名, 延べ 12 名 (すべて異なる)
  - 課題設定: 49 パターン
  - 発話数: 延べ 3568 発話 (延べ 445 課題対話)
- 第 2 回対話実験 (MAD2)
  - 日本語話者: 1 日あたり 1 名, 延べ 11 名 (すべて異なる)
  - 英語話者: 1 日あたり 1 名, 延べ 11 名 (すべて異なる)
  - 課題設定: 8 パターン
  - 発話数: 延べ 3404 発話 (延べ 69 課題対話)

表 1 発話に含まれる平均単語数
Table 1 Average number of words per utterance

	BTEC	SLDB	MAD1	MAD2
日本語	6.87	13.30	10.00	12.57
英語	5.87	11.27	10.25	11.06

表 2 発話に含まれる平均文数
Table 2 Average number of sentences per utterance

	BTEC	SLDB	MAD1	MAD2
日本語	1.07	1.35	1.29	1.44
英語	1.08	1.38	1.61	1.54

表 3 日本語における単文と複文の割合
Table 3 Simple and complex sentences in Japanese

	BTEC	SLDB	MAD1	MAD2
単文	82.8%	65.9%	68.3%	72.0%
複文	17.2%	34.1%	31.7%	28.0%

5. 議　　論

5.1 基本特性

対話実験により得られたデータの基本特性として, 発話に含まれる平均単語数, 発話に含まれる平均文数, 日本語における単文と複文の割合を調査した. BTEC,

☆ SLDB 収集時に経験的に得られた適切な値を採用した.

☆☆ 10 秒の音声を速やかに通訳しても音声で伝えるのに 10 秒要し, 相手が 10 秒で答えたものを再度通訳して音声で伝えるのに 10 秒要するとすれば, 合計 40 秒となる.

表 4　発話の分類
Table 4　Classification of utterances

	MAD1				MAD2			
	日本語		英語		日本語		英語	
a	518 発話	(29%)	440 発話	(24%)	359 発話	(22%)	292 発話	(17%)
c	74 発話	(4%)	48 発話	(3%)	75 発話	(5%)	93 発話	(5%)
e	877 発話	(50%)	878 発話	(49%)	910 発話	(55%)	940 発話	(54%)
ac	0 発話	(0%)	0 発話	(0%)	11 発話	(1%)	1 発話	(0%)
ae	302 発話	(17%)	428 発話	(24%)	265 発話	(16%)	422 発話	(24%)
ce	0 発話	(0%)	2 発話	(0%)	23 発話	(1%)	6 発話	(0%)
ace	0 発話	(0%)	1 発話	(0%)	3 発話	(0%)	1 発話	(0%)
no	0 発話	(0%)	0 発話	(0%)	1 発話	(0%)	2 発話	(0%)
合計	1771 発話	(100%)	1797 発話	(100%)	1647 発話	(100%)	1757 発話	(100%)

SLDB の値とともにその数値を表 1，表 2，表 3に示す．なお，日本語における単文と複文の分析手法は文献[9] による．

表 1，表 2，表3によれば，MAD の基本特性はSLDB に近く，BTEC はそれらと異なることがわかる．具体的には，BTEC は文が短く，単文が多い．

5.2　確 認 発 話

MAD の基本特性は SLDB に似ていることがわかったが，SLDB には認識誤りや翻訳誤りが含まれていないという違いがある．そこで，誤りに起因する確認発話の割合を調査した．まず，MAD データを次の三つに大きく分類してみた．

(a) 働きかけの発話
(c) 誤り等に起因する確認の発話
(e) それ以外の発話 (多くは回答等の情報伝達)

その具体的な作業手順を次に記す．

(1) 1 発話中で複数の意味を持つ場合は該当する二つ以上のマークを付与した．
(2) “May I help you?” に対する応答において「. . . したい」という趣旨の発話は次のように分類した．
  – 疑問文 → a
  – 肯定文 → e
  – 「. . . したいんだけど．」 → ae
(3) 「. . . お願いします」という文は，二泊・サラダ等目的語を指定している場合 e とした．それ以外は ae とした．

ラベル付与できないものを no とした上で得られた結果を表 4に示す．

今回の実験は音声認識性能が著しく良い場合の音声翻訳システムとみなせるが，そのような条件下で誤りに起因する確認発話の割合は 3 〜 5% であった．なお，この確認発話の内容は定型的な表現の確認発話と相手の発話内容を入れた確認発話の二つに分類することができる．それぞれの典型例を示す．

- 定型的な表現の確認発話
  – 「あ，もう一度言っていただけますか．」
  – “Could you repeat what you said? I couldn’t understand.”
  – “Sorry, can you repeat that?”
- 相手の発話内容を入れた確認発話
  – 「はい六月十日午前九時二十五分出発，あってますか．」
  – “Are you saying that I can have the single room for tonight, and then exchange it for the double room tomorrow?”
  – “Did you say you’d like to find a Chinese restaurant here?”

直前の相手発話の翻訳結果の意味がまったく把握できなかった場合に定型的な表現の確認発話がなされ，無理に解釈すれば一部の意味が把握できるような場合に相手の発話内容を入れた確認発話がなされるように見える．

BTEC には「もう一度言ってください．」というような例文が基本表現として含まれているため，定型的な表現の確認発話はおおむね BTEC とそれに基づく言い換え表現で扱うことができる．相手の発話内容を入れた確認発話については今後その取り扱いを検討する必要がある．

5.3　課題達成と発話数の関係

第 1 回対話実験シリーズで実現可能性の検証ができたので，第 2 回対話実験シリーズでは課題達成と発話数の関係を調べ，主に誤りに起因する確認発話が課題達成に及ぼす影響を調査することを試みた．分析対象とすべきサブタスクとして次の 12 種類を定めた．

- サブタスク 1: 参加するツアーとその日程を決め，確認する．
- サブタスク 2: ツアー参加のホテル送迎時間，方法等を確認する．
- サブタスク 3: 主な観光場所，ポイントを確認する．
- サブタスク 4: 支払額を確認する．

表 5　課題達成と発話数の関係
Table 5　Relationship between task achievement and the number of utterances

	ペア 1	ペア 2	ペア 3	ペア 4	ペア 5	ペア 6
サブタスク 1	9	15	14	9	9	4
サブタスク 2	5	7	12	—	7	6 (1)
サブタスク 3	11	38 (10)	27	18	24 (5)	14
サブタスク 4	5	12	7 (1)	11 (2)	5	5
サブタスク 5	2	2	3	4	2	—
サブタスク 6	2	3	9 (4)	4	3	4
サブタスク 7	2	2	4	3	2	8
サブタスク 8	3	7 (2)	6	5	4	3
サブタスク 9	9	14 (7)	8	9	4	9
サブタスク 10	3	3	3	5	2	4
サブタスク 11	20 (4)	3	4	3	5	5
サブタスク 12	3	8 (2)	5	4	4	6

- サブタスク 5: 予約者の名前を確認する.
- サブタスク 6: 予約者の連絡先を確認する.
- サブタスク 7: 食事の際に禁煙席に座る.
- サブタスク 8: 飲み物を注文する.
- サブタスク 9: 食事を注文する.
- サブタスク 10: 飲み物のお代わりをする.
- サブタスク 11: 日本料理の材料を説明する.
- サブタスク 12: 生魚が大丈夫か確認し，必要なら代用料理を準備する.

第 2 回対話実験 11 日間のうち，対話観察者 3 名の合意により，主観評価の良かった 3 組 (ペア 1 ～ 3) と，良くなかった 3 組 (ペア 4 ～ 6) について，サブタスクの達成に要した発話数を表 5に示す. 括弧内の数字が誤りに起因する発話のやり取り数を示す. — は対応するサブタスクがなされなかったことを示す.

表 5によれば，発話数が多くなるのは必ずしも誤りに起因する発話が多いためではない. むしろ，話者の個性に強く依存しているようである.

5.4　効率の良いコミュニケーションと豊かなコミュニケーション

課題達成に要する発話数は必ずしも誤りに起因する発話により増えるわけではなく，話者の個性に依存して変化するようなので，MAD データの内容を分析してみた. その結果，どうやら用件のみを短く言いたがる話者と，相手に配慮しながら，あるいは会話を楽しみながら話す話者がいることが観察できた. そこで，前者を効率の良いコミュニケーション，後者を豊かなコミュニケーションと名付けることにした. それぞれの典型例を次に示す.

- 効率の良いコミュニケーション
  - 「タクシー乗り場はどこですか. 」
  - 「私は黒のジャケットを探しています. 」
- 豊かなコミュニケーション
  - 「タクシーに乗りたいのですが，どこでタクシーに乗ればいいんですか. 」
  - 「春のジャケットを探しています. 青色のはありますか. 」

効率の良いコミュニケーションの話者は単文でしかも文を単位とする発話を好む傾向がある. 豊かなコミュニケーションの話者は複文あるいは複数の文を同時に発話することを好む傾向がある.

「タクシー乗り場はどこですか. 」という例文は BTEC に含まれる典型的な基本表現である. しかしながら，「タクシーに乗りたいのですが，どこでタクシーに乗ればいいんですか. 」というような複文のパターンは BTEC や BTEC に基づく言い換え表現には含まれていない. なぜならば，文末表現の細かい差異を除けば，「タクシー乗り場はどこですか. 」という基本表現から集められる言い換え表現は典型的には次のようなパターンのみだからである.

- 「タクシー乗り場はどこにありますか. 」
- 「タクシー乗り場はどこになりますか. 」
- 「タクシー乗り場を教えてください. 」
- 「どこに行けばタクシーに乗れますか. 」

今回の対話実験によれば，見知らぬ人に尋ねる場合，日本語話者のみならず英語話者でも相手に配慮した言い回しが頻度的にも多く好まれる傾向が見られた. 日本語話者と英語話者で課題が若干異なるが，そのような事例を次に示す.

- 「タクシーに乗りたいんですけど，乗り場を教えてください. 」
- 「タクシーに乗りたいのですが，どこに行けばいいですか. 」
- “I'd like to go downtown. Where can I catch a bus?”
- “I need to go downtown. Could you tell me

where I get on the bus?"

したがって，今後はたとえば言い換え表現収集で利用する例文の種類を変えることなどにより，実際に好まれる表現を収集することが必要となる．

5.5 今後の展望

豊かなコミュニケーションの話者には，実験後のアンケートにおいて10秒よりもっと長く一度に話せる方が話しやすいというコメントを寄せている人もいた．したがって，そのような態度の人に効率の良いコミュニケーションを強要するのは好ましくないかもしれない．話者適応等の技術により音声認識性能がどの程度向上するか確認した後で対策を講じる予定である．

言語モデル的にはBTECとSLDBを組合わせることでMADデータに適したものが構築できる可能性が期待できるので，そのような適応技術の研究を進める予定である．

また，誤りに起因した確認発話の情報を用いれば，その直前の翻訳結果と話者の言語行動の関係を分析することもできるので，そのような対話の特性に着目した翻訳技術評価の研究も今後の検討課題である．

第2回対話実験シリーズでは英語の合成音声の品質が向上した．その結果，英語話者の一部には画面表示に頼らず対話を進める人もいた．日本語のわからない英語ネイティブ話者であるものの，日本に居住している点に注意を要するが，合成音声の品質向上も円滑な対話進行を実現する上で重要な課題である．

6. むすび

コーパスベース音声翻訳研究のために機械を介在した対話形式による日本語話者と英語話者のコミュニケーションデータを収集した．引き続き対話実験結果のさらなる分析を行う予定である．また，翻訳評価や音声認識評価のためのテストセットを選定し，それら要素技術の評価実験に利用する準備を始めている．

今後は音響処理に重点を置いた対話データ収集を実施した後，タイピストの代わりに音声認識器を導入し，フィールドデータを収集する計画である．

謝辞 対話データ収集実験を実施するにあたり貢献いただいた高嶌浩司，松井孝典，染川智子 各氏に心より感謝申し上げる．

本研究は通信・放送機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものである．

参考文献

1) 菅谷史昭, 竹澤寿幸, 隅田英一郎, 匂坂芳典, 山本誠一: 音声翻訳システム: ATR-MATRIXの開発と評価, 情報処理学会論文誌, Vol. 43, No. 7, pp. 2230–2241 (2002).
2) Takezawa, T., Sumita, E., Sugaya, F., Yamamoto, H. and Yamamoto, S.: Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world, *Proc. 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Vol. I, pp. 147–152 (2002).
3) 金城由美子, 青野邦生, 安田圭志, 竹澤寿幸, 菊井玄一郎: 旅行会話基本表現に対する日本語パラフレーズデータの収集, 言語処理学会第*9*回年次大会発表論文集 (2003).
4) 竹澤寿幸, 中村篤, 隅田英一郎: ATRの会話音声翻訳研究用データベース, 音声研究, Vol. 4, No. 2, pp. 16–23 (2000).
5) Takezawa, T., Sugaya, F. and Kikui, G.: Using bilingual conversational expressions in speech translation, *Proc. Linguistics and Phonetics 2002 (LP 2002)* (to be published).
6) 古瀬蔵, 山本和英, 山田節夫: 構成素境界解析を用いた多言語話し言葉翻訳, 自然言語処理, Vol. 6, No. 5, pp. 63–91 (1999).
7) Sumita, E.: Example-based machine translation using DP-matching between word sequences, *Proc. ACL-2001 Workshop on Data-Driven Methods in Machine Translation*, pp. 1–8 (2001).
8) Campbell, N.: CHATR: A high-definition speech re-sequencing system, *Proc. ASA/ASJ Joint Meeting*, pp. 1223–1228 (1996).
9) 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝: 節境界自動検出ルールの作成と評価, 言語処理学会第*9*回年次大会発表論文集 (2003).