2004－SLP－50　(13)
2004／2／7

翻訳システムを介した音声対話における教示と言語表現の関係

竹澤 寿幸† 菊井 玄一郎†

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．特に話者に与える教示により，実際の言語表現がどのように変わるか調査した．大きめの声で明瞭に10秒以内で話すというものと，それに加えて伝達したい情報を分割し簡潔に話すという二つのものを試みたところ，日本語話者の半分，英語話者の全員が短く話すことができた．さらに，日本語の単文の比率は約70%から約80%に増えた．実験の概要を報告し，音声翻訳性能への影響を議論する．

A Study on Instructions and Linguistic Expressions in Machine-Translation-Aided Bilingual Spoken Dialogues

Toshiyuki Takezawa† and Genichiro Kikui†

A large bilingual corpus of English and Japanese is being built at ATR Spoken Language Translation Research Laboratories in order to improve speech translation technology to the level that people can use a portable translation system for traveling abroad, dining and shopping, and hotel situations. As a part of these corpus construction activities, we have been collecting spoken dialogue data using an experimental translation system between English and Japanese. We adopted two different instructions. One, Instruction A, is that utterances must be completed within ten seconds. The other, Instruction B, is that subjects should speak briefly and concisely. Half of the Japanese speakers were able to make their utterances short, while all of the English speakers were able to make their utterances short. As for the percentage of simple and complex sentences in Japanese, approximately 70% of the sentences for Instruction A were simple sentences, while approximately 80% of the sentences for Instruction B were simple sentences. We present a detailed discussion to explain these numerical data.

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大規模でかつ実際的なコーパスが必要である．世界的にもコーパスベースの音声翻訳技術が注目を集めているところであり，ヨーロッパではTC-STAR (Technology and corpora for speech-to-speech translation)[2]，米国ではDARPA Babylon[3]という大規模なプロジェクトが近年開始されている．

大規模なコーパスとして，ATRでは旅行会話基本表現集BTEC (Basic Travel Expression Corpus)[4]の構築を進めている．日本人が海外を旅行したり，あるいは，外国人が日本を旅行したりする際に遭遇する様々な場面の表現を日英対訳で収集し，20万文以上の規模となっている．その一部は中国語等，他の複数の言語にも翻訳中である[5]．

さらに，実際的なデータ収集として，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るコーパスMAD (Machine-translation-Aided bilingual spoken Dialogue corpus)の構築も進めている[6]．その第1段階として，機械翻訳部に重点を置き，音声認識部の代わりに人間(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ており，1万発話を越える規模となっている[7]．

そのような実際的な対話データ収集の一環として，話者に与える教示により，実際の言語表現がどのように変わるか調査することを目的に実験を行った．具体的には，大きめの声で明瞭に10秒以内で話すというものと，それに加えて伝達したい情報を分割し簡潔に話すと

† (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

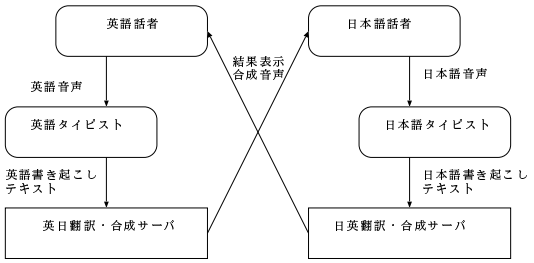

図1 実験システム構成
Fig. 1 Experimental system configuration

いうもの等を試みた．本稿では，実験の概要を報告し，音声翻訳性能への影響を議論する．

2.では実験システム構成を述べる． **3.**では実験の概要を述べる． **4.**では収集された対話データの基本特性を報告し，議論を行う．最後に **5.**で全体をまとめる．

2. 実験システム構成

今回の実際的な対話データ収集においても，機械翻訳部に重点を置くこととし，音声認識部の代わりに人間(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主に採用した．

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)[8] 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)[9] 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[10] を使用した．英語音声合成には AT&T Labs' Natural Voices^{TM} を使用した．

なお，一部の対話データ収集の際に日本語音声認識を行った．それには音声翻訳システム ATR-MATRIX[1] の日本語音声認識部 SPREC[11] を拡張したものを使用した．実験システムとしては，図 1の日本語タイピストの代わりに日本語音声認識が入ることになる．

☆ 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．

3. 対話データ収集

3.1 話者への教示

これまでの MAD コーパス収集[6),7)] では，通訳者を介したバイリンガル旅行会話コーパス SLDB[12] 収集時に経験的に得られた知見をもとに，通訳者ではなく翻訳システムが介在することに起因する項目を付加した教示を話者に与えていた．具体的には，実験参加者である話者に対して，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．

(1) 大きめの声で明瞭に話す．
(2) 1 回の発話は 10 秒以内とする☆☆．
(3) 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．
(4) 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．

(1) と (2) が SLDB 収集時の知見に基づくものであり，(3) と (4) が翻訳システムが介在することに起因する項目である．このような話者への教示を採用した理由は二つある．一つは収集された対話データの基本特性を SLDB と比較・議論することが可能と期待できるからである．もう一つは話者に過度の負担をかけないものが適切と判断したため，必要最小限にとどめたからである．今回の対話データ収集においても，この従来のものを基準に選び，これを「インストラクション A」と名付けることにする．

音声翻訳システム ATR-MATRIX 開発時の知見によれば，音声認識部，機械翻訳部ともに文長が短い方が性能が良い傾向がある[1)] ことから，話者への教示を増やしたものを試みることにした．具体的には，インストラクション A に次の項目を加えたものを実施した．

(5) 伝達したい情報を分割して，短く簡潔に話す．
(6) 「わたしは」「わたしの」「あなたに」等の語句を加えてみると，英語への翻訳がうまくいくことがある．

(5) が短く話すように誘導するものである． (6) は特に日英方向の機械翻訳を意識したものである．これを「インストラクション B」と名付けることにする．

さらに，音声認識を導入する際の話者への教示の試みとして，インストラクション B に次の項目を加えたものを実施した．

(7) なるべく抑揚を付けずに話す．
(8) 一定のリズムで話す．

☆☆ SLDB 収集時に経験的に得られた適切な値を採用した．

表1 話者への教示スケジュール
Table 1 Schedule of instructions to users

	前半 6 日間	後半 6 日間
午前	AT	BT
午後	BT	CT, CR

これを「インストラクション C」と名付けることにする．

実際の機器操作としては，ボタンを押してから話者に話させ，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．なお，インストラクション C の一部で，日本語話者についてのみ実際に音声認識装置を使ったデータも集めた．区別するために，タイピスト方式を T，音声認識を使った場合を R という記号を付加すれば，データとしての区別は 4 種類となり，それぞれ AT, BT, CT, CR とラベル付与することにする．

3.2 実験の実施

処理速度としては 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，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☆．

ユーザインタフェースとしては，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を採用した．このようにして収集した量は次の通りである．第 4 回めのデータ収集に相当するので，MAD4 と呼ぶことにする．

- 名称: MAD4
- 実施時期: 2003 年 6 月から 7 月の 12 日間
- 日本語話者: 12 名 (日替わり)
- 英語話者: 12 名 (日替わり)
- 課題設定: 16 パターン
- 発話数: 延べ 3666 発話 (延べ 166 課題対話)

話者への教示は，前半 6 日間については，午前にインストラクション A，午後にインストラクション B を実施した．対話の課題は午前と午後のものを毎日入れ替えて，バランスさせた．後半 6 日間については，午前にインストラクション B，午後にインストラクション C を実施した．対話の課題は前半同様に午前と午後のものを毎日入れ替えて，バランスさせた．話者への教示スケジュールを表 1に示す．

☆ 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．

4. 議　　論

4.1 全体の基本特性

得られた対話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 の値とともにその数値を表 2，表 3，表 4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[13] による．

表 2，表 3，表 4によれば，MAD4/AT の基本特性は，英語の発話に含まれる平均文数が増えている点を除けば，おおむね SLDB に近いといえる．その理由は話者への教示が SLDB の際のものを基礎にしているためと考えられる．なお，英語の発話に含まれる平均文数が増えているのは，一部の英語話者が返事をする際に “okay” や “yes” を発話の最初にほぼ常に加えていたためであった．

MAD4/AT と MAD4/BT を比較すると，発話に含まれる平均単語数は，日本語で 88%, 英語で 76% に減っており，意識して話すことにより短く話せることがわかった．同時に日本語の単文と複文の比率を比べると，MAD4/AT で全体の 70% が単文であったものが，MAD4/BT では全体の 80% が単文であり，10 ポイントの増加が見られる．MAD4/BT と MAD4/CT では大きな違いは見られないが，日本語側のみ音声認識を利用した場合である MAD4/CR では発話がさらに短くなる傾向がある．

4.2 話者による違い

全体の平均的な観点から MAD4/BT は MAD4/AT に比べて短くなっていることがわかった．そこで，次に話者による違いを調査した．前半 6 日間について，日本語話者を順に J1, J2, J3, J4, J5, J6, 英語話者を順に E1, E2, E3, E4, E5, E6 とし，話者別に求めた発話に含まれる平均単語数と，変化率を表 5に示す．

表 5によれば，日本語話者の半分，英語話者の全員が教示に従い，発話を短くすることができたことがわかる．この違いの理由としては，日米によるオーラル教育の差に起因する可能性がある．ただし，日本で集められる英語話者は英会話の指導者が多く，言語運用能力にバイアスがかかっている可能性もある．

表2　発話に含まれる平均単語数
Table 2　Average number of words per utterance

	BTEC	SLDB	MAD4/AT	MAD4/BT	MAD4/CT	MAD4/CR
日本語	6.87	13.30	11.13	9.78	9.01	8.02
英　語	5.87	11.27	12.60	9.56	9.54	8.97

表3　発話に含まれる平均文数
Table 3　Average number of sentences per utterance

	BTEC	SLDB	MAD4/AT	MAD4/BT	MAD4/CT	MAD4/CR
日本語	1.07	1.35	1.35	1.41	1.33	1.23
英　語	1.08	1.38	2.19	1.78	1.84	1.74

表4　日本語における単文と複文の割合
Table 4　Simple and complex sentences in Japanese

	BTEC	SLDB	MAD4/AT	MAD4/BT	MAD4/CT	MAD4/CR
単文	82.8%	65.9%	69.5%	79.8%	79.9%	83.5%
複文	17.2%	34.1%	30.5%	20.2%	20.1%	16.5%

表5　話者毎の発話に含まれる平均単語数
Table 5　Average number of words per utterance for each speaker

話者	MAD4/AT	MAD4/BT	変化率
J1	13.0	10.4	80.0%
J2	9.6	9.6	100.0%
J3	12.1	10.2	84.3%
J4	10.8	11.5	106.5%
J5	11.1	8.9	80.2%
J6	10.9	10.6	97.2%
E1	9.0	8.0	88.9%
E2	13.3	10.4	78.2%
E3	11.5	9.6	83.5%
E4	15.8	13.4	84.8%
E5	13.8	8.4	60.9%
E6	11.5	9.7	84.3%

4.3　発話のエントロピーと長さの分布

さらに，MAD4/ATとMAD4/BTについて，各発話毎のエントロピーと長さの関係を調査した．MAD4/ATの結果を図2に，MAD4/BTの結果を図3に示す．それぞれ横軸が発話を単位とするエントロピー，縦軸が発話長つまり発話に含まれる語数を示す．BTECで訓練した言語モデルにより，エントロピーを求めた．日本語発話を対象として，グラフ化したものである．

図2と図3を比較すると，与える課題をバランスさせたためか，発話を単位とするエントロピーの分布に大きな違いは見られないが，MAD4/ATには長い発話が見られるのに対し，MAD4/BTでは長い発話が減っていることがわかる．

表6　代名詞等の文を基準とした出現率
Table 6　Occurrence of pronouns per sentence

語句	MAD4/AT	MAD4/BT	変化率
わたし	0.99%	4.58%	462.6%
わたくし	0.00%	0.00%	—
あなた	0.00%	2.14%	∞
そちら	0.59%	0.15%	25.4%
こちら	0.79%	1.83%	231.6%
お客様	0.00%	0.15%	∞

4.4　代名詞等の出現頻度

インストラクションBでは，日本語話者に対して「わたし」「あなた」等の代名詞等を加えてみるような指示も行った．そこで，特に「わたし」「わたくし」「あなた」「そちら」「こちら」「お客様」の六つの語句に着目して，文を基準とした出現率を求めた．結果を表6に示す．

表6によれば，「わたし」の出現率はインストラクションにより4.6倍に増えているが，文を基準とすれば，もともと約1%の文で「わたし」が使われていたものが約5%の文で使われるようになったに過ぎず，依然として約95%の文では「わたし」が使われることはない．機械翻訳機能付きのウェブを介したソフトウェア開発[14]の知見によれば，翻訳の性能向上のために日本語の主語を不自然なまでに徹底的に補った文章も使われる

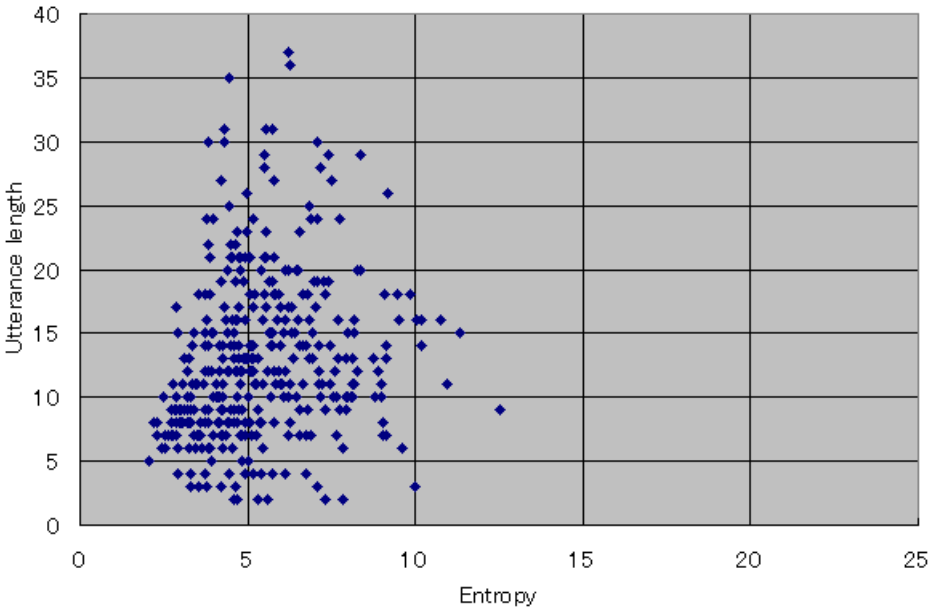

図2　発話のエントロピーと長さの関係 (MAD4/AT)
Fig. 2　Relationship between entropy of utterance and utterance length (MAD4/AT)

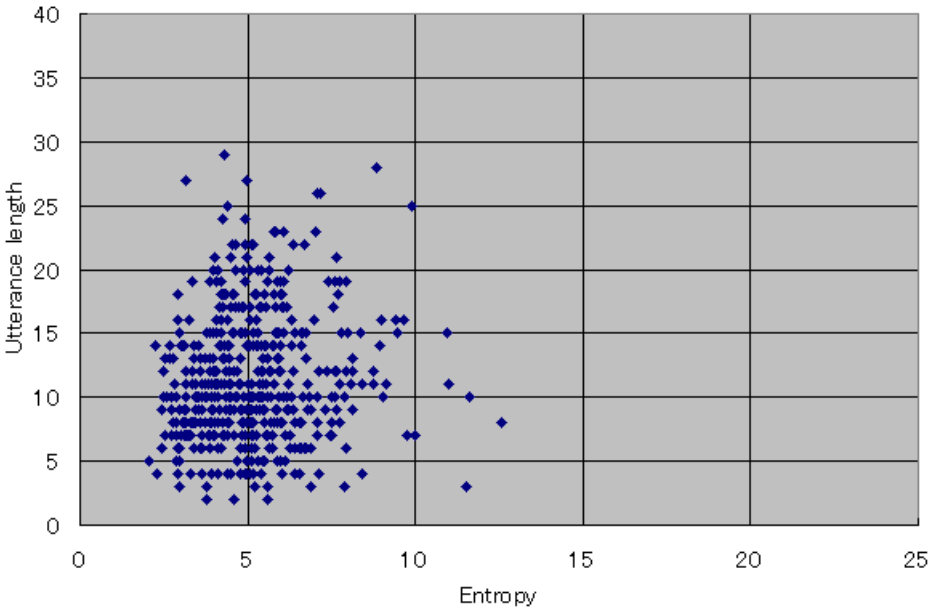

図3　発話のエントロピーと長さの関係 (MAD4/BT)
Fig. 3　Relationship between entropy of utterance and utterance length (MAD4/BT)

ようになるようであるが，対面対話で話し言葉の場合には代名詞等を補うのは難しいような結果である．

4.5　BTEC によるカバー率

ATR では大規模なコーパスとして BTEC の構築を進めている．実際的なコーパスである MAD と BTEC の共通部分が多ければ，音声翻訳性能の向上が期待できる．そこで，BTEC による MAD4/AT, MAD4/BT のカバー率を求めた．文を単位として完全一致した割合

表7 BTECによるカバー率
Table 7 Coverage based on BTEC

	MAD4/AT	MAD4/BT
全体	31.4%	35.0%
単文のみ	43.2%	42.0%

ではなく，類似した文が含まれる割合をカバー率とみなした．文長が0.8倍から1.2倍であり，その範囲で全体の80%以上の語が一致した場合を類似文と定義することにした．日本語について調べた結果を表7に示す．

表7を見ると，MAD4/ATで全体の70%が単文であったものが，MAD4/BTでは全体の80%に増えていることにより，全体的にはカバー率が増加している．しかし，単文のみで比較すると，ほぼ同じであり，むしろMAD4/BTは若干減少している．その理由は「わたし」「あなた」等の代名詞等を加えるような教示の影響と考えられる．

5. むすび

コーパスベース音声翻訳研究のために翻訳システムを介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更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，それら要素技術および統合技術の評価実験に利用する準備を始めている．

今後は，発話者の元の音声が聞こえる影響を調べるための実験を実施した後，タイピストの代わりに音声認識装置を導入し，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鈴木弥生，西野敦士，高嶌浩司，染川智子，伊藤玄，水町光徳 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) Höge, H.: Project proposal TC-STAR: Make speech to speech translation real, *Proc. 3rd International Conference on Language Resources and Evaluation (LREC)*, Vol. I, pp. 136–141 (2002).
3) Special Session: Multilingual Speech-to-Speech Translation, *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH)*, Vol. 1 (2003).
4) 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)*, Vol. I, pp. 147–152 (2002).
5) Kikui, G., Sumita, E., Takezawa, T., and Yamamoto, S.: Creating corpora for speech-to-speech translation, *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH)*, Vol. 1, pp. 381–384 (2003).
6) Takezawa, T., and Kikui, G.: Collecting machine-translation-aided bilingual dialogues for corpus-based speech translation, *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH)*, Vol. 4, pp. 2757–2760 (2003).
7) 竹澤寿幸, 西野敦士, 高嶌浩司, 松井孝典, 菊井玄一郎: 機械翻訳を介した対話データ収集のための実験システム, 情報科学技術フォーラム *(FIT)*, E-036, 一般講演論文集第 *2* 分冊, pp. 161–162 (2003).
8) 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).
9) 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).
10) Campbell, N.: CHATR: A high-definition speech re-sequencing system, *Proc. ASA/ASJ Joint Meeting*, pp. 1223–1228 (1996).
11) 内藤正樹, 山本博史, シンガーハラルド, 中嶋秀治, 中村篤, 匂坂芳典: 対話音声を対象とした連続音声認識システムの試作と評価, 電子情報通信学会論文誌 *D-II*, Vol. J84-D-II, No. 1, pp. 31–40 (2001).
12) 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).
13) 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, pp. 517–520 (2003).
14) 野村早恵子, 石田亨, 船越要, 安岡美佳, 山下直美: アジアにおける異文化コラボレーション実験2002: 機械翻訳を介したソフトウェア開発, 情報処理, Vol. 44, No. 5, pp. 503–511 (2003).