対話の時間構造に着目した聞き上手な留守番電話の設計

Design of skillful-listening answering machine based on temporal structure of speech dialog

小島 敬 [1] * 川嶋 宏彰 [1] 平山 高嗣 [1] 松山 隆司 [1]
Kei Kojima[1] Hiroaki Kawashima Takatsugu Hirayama Takashi Matsuyama

[1] 京都大学大学院情報学研究科
[1] Graduate School of Informatics, Kyoto University

Abstract: In this research, a dialog system, which gives the impression of being a good listener, has been designed and the characteristic points of good listening are identified. We propose an approach of dividing dialogues into two modes, namely in prompting modes and understanding modes. In order to show the usefulness of this approach, we focussed on the situation given while using an answering machine, where the speaker has a motivation to talk but has not yet organized the topics, and have designed a response system, which motivates the user to speak. As a result the effectiveness of the combination of the prompting modes with the understanding modes, which give the speaker an opportunity to start to speak and the interpretation of the system's voice as feedback by the speaker has been verified as a mean to initiate an uttterance of the speaker.

1 はじめに

聞き手の上手な対話の運び方が，コミュニケーションスキルの一つとして捉えられるようになってきている．話し手は聞き上手な相手との対話により，聞き手に対して安心感を感じることができ，より多くの情報を負担無く提示することができる．対話システムにおいても，このような聞き手の立場は重要である．対話システムが，人に対して安心感を与える聞き方をすることにより，システムのリピータの増加であったり，人からの話をより多く引き出すことに繋がるからである．従来の対話システムは，タスク達成に重点がおかれていたのに対し，最近では聞き手としての役割に重点をおかれたシステムが提案されはじめている [3][4]．これらの研究については後述する．本研究では，この流れをさらに進め，対話システムが聞き上手になるための一つの設計指針として，話し手の意識に応じた聞き手の対話の運び方に着目し，システムの応答を設計する．

本研究での具体的な対話状況として，留守番電話での録音状況に焦点を当てる．その理由としては，ユーザがある程度話したい内容を持っているにも関わらず，聞き手としての機能が備わっていないため，安心感を持ってメッセージを残すことがしばしば困難となるからである．松尾 [1] は留守番電話の使いにくさの原因としてインタラクティブ性の欠如を挙げている．システムが聞き手としてのインタラクティブなやり取りを補えるよう応答を設計することで，システムの聞き方に対する知見が得られると考えられる．

留守番電話という状況とは異なるものの，聞き手のインタラクティブ性に注目した対話システムが近年注目されている．小林ら [3] は，話し手の韻律情報に適応した応答を行うシステムを実装している．竹内ら [4] は，人と人の対話データを学習し，決定木により適切な応答タイミングを決定するシステムを構築している．これらの研究では，システムが人との対話において相槌を生成し，対話のインタラクティブ性を実現している．このような相槌応答を留守番電話に用いた先駆的研究として，向後ら [5] はユーザの発話に合わせて相槌を打つシステムを設計している．しかしながら，このシステムは話し手が急な相槌に対応できず，相槌が逆効果になるという結果になっている．これは，従来の留守番電話のような聞き手がいないイメージで発話を行っている際に，聞き手から相槌が挿入され，相槌が雑音として解釈されることが原因として考えられる．そこで本研究では，システムが話し手に応答することの重要性に加え，話し手が発話しやすい状態を作るための，システムから話し手への働きかけについて考える．つまり，システムからの働きかけを冒頭に加えることで，聞き手としてフィードバックを送る準備が整っていることを表現できる．我々は，

* 京都大学大学院情報学研究科
606-8501 京都府京都市左京区吉田本町
kojima@vision.kuee.kyoto-u.ac.jp

聞き手から情報を受け取ろうという態度を示すモードから話し手に理解表現を行うモードへの流れを作ることを，聞く行為として捉える．本研究では，このような流れを中心に，留守番電話での応答を設計し，話し手に安心感を与える対話をシステムが作ることができるかを検証する．

以下 2 章において，聞く行為の詳細を述べ，3 章で留守番電話の具体的な応答の流れを設計する．4 章で対話実験による有効性を検証し，5 章で結論を述べる．

2 聞く行為の流れ

2.1 聞き手はどのように対話を運ぶべきか

話し手から聞き手に情報を伝える場面においては，聞き手は，話し手に対して情報を受け取る準備ができていることを表現しなければならない．対面対話においては，対話の開始時に，聞き手が話し手以外の相手と話していたり，別の作業を行っていると，話し手は話し始めるきっかけを失ってしまう．電話対話の場合においては視覚情報が無いため，対面対話に比べて，聞き手側から積極的に，情報を受け取ろうという態度を示す必要がある．このような態度を，聞き手の「促しモード」と呼ぶこととする．話し手は聞き手からの「促しモード」を通じて，聞き手に情報を伝達できるようになる．話し手の話が開始されると，聞き手は話し手に対して情報取得に成功していることを表現する．このような聞き手の態度を「理解モード」と呼ぶ．聞き手の「理解モード」により，話し手は聞き手の反応を窺いながら徐々に話を進めていくことができる．また，安心感を持って対話を終了することができる．次節では，話し手の対話中での意識と対応させて聞き手のモードについて述べる．

2.2 聞く行為の流れ

2.1 節での考え方をふまえ，聞き手による対話中のモードについて，話し手の意識と対応させて整理する．まず，話し手の対話意識は以下のような 3 つのモードを持つと考える．

1. Disconnected モード：聞き手の存在を感じられない状態（フィードバックを期待できない）
2. Connected モード：聞き手が話し手（自分）に対して意識を向けていると感じる状態（フィードバックを期待する）
3. Accepted モード：聞き手が話し手（自分）の発話内容を受け止め，理解していると感じる状態

聞き手の対話姿勢は，話し手のモードを遷移させるよう働きかけることで，先に挙げた 2 つのモードを持つものとする．

1. 促しモード：話を聞こうとする態度を表現する．
2. 理解モード：話し手に対して情報取得，理解を表現する．

■聞き手の促しモードの役割： 促しモードの役割は，話し手に対して聞き手が情報を取得する準備ができていることを表現することである．つまり，話し手の対話意識を Disconnected モードから Connected モードへと遷移させる働きである．具体的には，対話の序盤に聞き手側から話し手に対して簡単な働きかけを行う．このような働きかけにより，聞き手からのフィードバックを期待する．

■聞き手の理解モードの役割： ここでの機能は，話し手の対話意識を Connected モード→ Accepted モードへと遷移させる．そのために，聞き手は話し手に対して理解表現を行う．ここでの具体的な応答方法については 3 章で述べる．

以上をまとめると，聞き手は対話の始めに促しモードにより，話し手の対話意識を Disconnected モードから Connected モードへと遷移させ，話し手対して話しやすい状況を作る．話し手の発話が開始されると，聞き手は話し手に対して理解表現を行い，話し手の意識を Connected モードから Accepted モードへと遷移させる．次章では，このような流れに基づいて，留守番電話の応答について具体的に設計を行う．

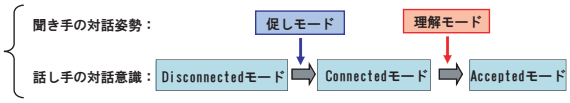

図 1 話し手と聞き手のモードの関係

3 留守番電話の応答設計

3.1 留守番電話の場面

本章では，留守番電話の応答について，前章での聞き手の対話姿勢に基づいて設計を行う．

留守番電話においては，話者同士の関係により発話スタイルが大きく変わることが考えられるので，対話状況をある程度絞る必要がある．本研究においては，友人や家族などの私的な場面ではなく，仕事関係など公的な場面を想定する．その理由としては，私的な関係で電話をかける動機は，相手と話すことが目的である場合が多く，メッセージを残す必要が無いからである．一方で公的な関係で電話をかける場合は，いい加減な切断が許されにくいことや，相手に伝える必要がある用件を持っている場合が多いからである．従って本研究では，留守番

電話の対話状況として，研究室に電話をかける公的な状況を想定した応答をデザインする．

3.2 応答の流れ

本章では 2 章で述べた対話の流れに基づいて，留守番電話での応答の流れを以下の 3 つの対話フェーズで構成する．ここからは電話をかけたユーザを話し手，電話がかかってきたユーザをホストユーザ，システムを聞き手と呼ぶ．

1. **導入部**…話し手の基本情報を聞くとともに，ホストユーザが電話に出れないことを聞き手が伝える部分
2. **メッセージ入力部**…話し手がメッセージを伝える部分．聞き手は話し手の用件を聞きながら相槌を打つ．
3. **クロージング部**…話し手のメッセージを受けたことを聞き手が伝える部分

ここでの対話フェーズは，2 章での聞き手の対話姿勢に基づいた流れを汲んでいる．つまり，対話の冒頭部分に聞き手として機能を果たすことを示し，本題を聞く準備を整える．話し手から本題を聞いている時は，聞き手から話し手に理解表現を行う，という流れである．それぞれの対話フェーズについての詳細は以下の通りである．

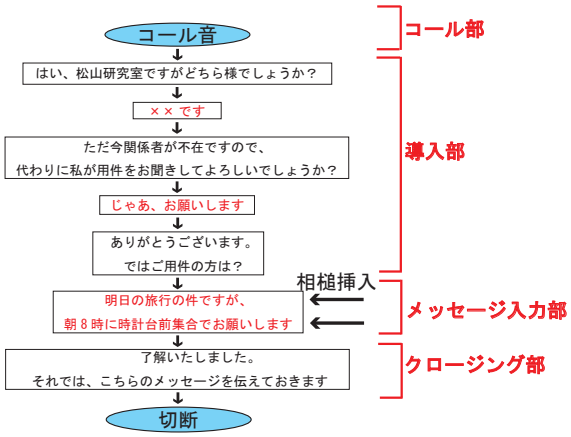

図 2　具体的な音声応答の流れ

■導入部の役割：　導入部は，聞き手側から話し手に対して積極的にやり取りを行う部分である．この部分は促しモードの役割を持ち，話し手の意識を Disconnected モードから Connected モードへと意識を遷移させる役割を持つ．向後ら [5] の留守番電話に相槌を挿入するシステムにおいて，急な相槌に話し手が対応できないという結果となったのは，従来の留守番電話の応答を受けて話し手が Disconnected モードで発話を行っている時に，急に聞き手が現れ，話し手が相槌に対応できないということが考えられる．つまり，導入部と相槌を組み合わせることにより，話し手が Connected モードで発話するため，相槌の雑音性が無くなると考えられる．

■メッセージ入力部の役割：　メッセージ入力部は本題を話し手が発話する部分である．聞き手は理解モードの役割を果たし，話し手の発話に対する理解表現を行う．これによって，話し手が Connected モードからさらに Accepted モードへと遷移することを期待する．具体的な応答としては，話し手の発話を受けて聞き手が相槌を挿入する．では具体的にどのような相槌を使うべきであろうか．一言で相槌といっても，様々な表現のものが考えられる．どのような相槌をとるべきかについては 3.3 節にて考察する．

■クロージング部の役割：　クロージング部は理解モードの役割を持ち，話し手の話全体に対して理解できたこと，対話を終了することを伝える．そうすることで，話し手の対話意識を Accepted モードへと遷移させる．

以上の流れは, 話し手の対話意識を Disconnected モードから Accepted モードへと遷移させ，話し手に安心感を持って対話を進めてもらうことが狙いである．次節ではメッセージ入力部でのシステム側の相槌について，どのような相槌が望ましいかを分析および考察し，4 章では対話実験により提案した応答の有効性を検証する．

3.3 メッセージ入力部における相槌

メイナード [6] は相槌の機能について, 以下のように分類している．

1. 続けてというシグナル
2. 内容理解を示す表現
3. 話し手の判断を支持する表現
4. 相手の意見, 考え方に賛成の意志表示をする表現
5. 感情を強く出す表現
6. 情報の追加, 訂正, 要求などをする表現

このうち，公的な対話状況で，話し手に対する情報取得表現となるものは，1 と 2 が主であると考えられる．本研究では 1 と 2 の相槌の違いとして，話し手に対する理解度の違いに注目する．つまり本節では，これらの相槌について，1 を「促しの相槌」（理解モードとしての働きが弱く，発話を促す機能が強い），2 を「理解の相槌」（理解モードとしての働きが強く，発話を促す機能が弱い）と捉え，それぞれの相槌がどのような発話特徴であるか，どのように使い分けられているかを検証する．

3.3.1 相槌の機能と発話特徴の関係

本節では対話コーパスを基に，理解モードの働きの弱い「促しの相槌」と，理解モードの働きの強い「理解の相槌」の発話特徴の違いを分析する．分析に用いたコーパスは国立情報学研究所が公開している RWCP-SP96 音声データベース（96 年版）[7] を用いた（店員と顧客とのフォーマルな対話）．

分析方法は，対話の冒頭部分（聞き手である店員は 4 名，店員 1 名あたり 12 対話，1 対話 1〜2 分程度）での店員の相槌音声のみを取り出し，被験者にそれぞれの相槌音声のラベルを付けてもらった．相槌のラベルは，それぞれの相槌が 1:促しの相槌（続けてという表現であるか），2:どちらともとれる，3:理解の相槌（話し手に対する理解を示しているか），の 3 パターンであり，合計 5 人の被験者がラベル付けを行った．相槌の分類については，5 人中 4 人が 1，残り 1 人が 2 か 1 と答えたものを促しの相槌，5 人中 4 人が 3，残り 1 人が 2 か 3 と答えたものを理解の相槌として，それぞれの相槌を分類した．

コーパス中での聞き手の相槌の言語情報は，「あはい」，「はい」，「ええ」，「はあ」の 4 種類であった．促しと捉えた人数が多い相槌は「はい」，「ええ」という音声で，理解と捉えた人数が多い相槌は，「あはい」という音声であった．それぞれ分類された相槌を，発話長，F0 平均，F0 回帰係数（発話終端 50ms）を計算し，得られたパラメータについて有意水準 5% で t 検定を行った．その結果，発話長において有意差があることが検証された．これは，「はい」，「ええ」と「あはい」という言語的特徴の違いが発話長の違いとして現れたと考えれらる．しかしながら，言語情報として同じ「あはい」であっても「あ」と「はい」の間のポーズが短いものが促しとして解釈されていたものもあったので，「あ」と「はい」の間のポーズ時間によって，理解度が変わることも示唆される．

3.3.2 相槌が挿入される先行話者の発話特徴

次に対話中で，聞き手が二つの相槌を使い分ける要因を，先行話者の発話特徴から分析する．

電話対話における話し手と聞き手の対話データ（8 名の話し手が用件を伝える場面，1 名あたり 4 対話収録）から，話し手の音声のみを取り出し，ポーズに挟まれた部分（閾値 30dB 以下）の発話を切り出した．切り出した音声に対して，発話特徴から分類された促しの相槌（言語情報が「はい」），理解の相槌（言語情報が「あはい」）を 300msec のポーズを挟んで挿入し，それぞれ 1 つの評価音声とした．被験者には，それぞれの評価音声を聞いてもらい，挿入された相槌が良いか悪いか（1（悪い）から 6（良い）の 6 段階）を被験者に判断してもらった．分割した発話数は合計で 251 発話で，促し，理解の相槌を同一の発話に対して組み合わせたので，合計で 502 個の音声を聞き取ってもらった．また，提示する順番に関しては，ランダムに提示せず前後関係を把握してもらうために，順序関係を保ったまま合計 5 名の被験者に評価してもらった．

決定木（データマイニングツール Weka3-2[8] を利用）を用いて相槌を識別し，識別に大きな影響を与える要素を検証した．入力する情報としては，先行発話者の韻律情報（F0，パワーの平均値，終端部分の回帰係数），言語情報（節境界の切れ目の強さ，品詞情報），対話全体での発話の位置（ポーズを挟んで何番目に発話されたものか）とした．出力情報としては「相槌を行わない」（被験者の評価結果が促しの相槌，理解の相槌共に評価が 3 以下となっている場合），「理解の相槌を行う」（評価が 3 より大きく，促しの相槌よりも評価値の高いもの），「促しの相槌を行う」（上記以外の場合）という 3 値とした．

決定木による分析の結果，識別に主に現れた要素としては，以下の情報が強く現れた．

- 各発話の節の切れ目の強さ
- 各発話が対話開始から何番目に発話されたものか

決定木の葉に関して，決定木の識別率は，全体の発話のうち 84.4 %であった．

3.3.3 相槌に関する考察

促しの相槌，理解の相槌に関して，フォーマルな状況においては，「はい」に「あ」という気づきの音声が加わると，理解度が増して聞こえる表現となることが検証された．このような相槌は相槌全般からすると一つの例でしか無いが，本研究で注目する点は，これらの機能の違う相槌を意図的に使い分けることで，話し手の意識にどのような影響を与えるかという点である．

また，それぞれの相槌が挿入される箇所については，全般的に促しの相槌は 300msec のポーズ後で，言い淀みやフィラーの直後以外に挿入されれば，悪い評価となることは少ない．理解の相槌は，発話が終盤に進む程，話し手の発話が比較的強めの節 [9] の切れ目（絶対境界，強境界）となる程，好まれるという結果であった．このことから，聞き手は意味の切れ目や，発話終了を窺いながら，徐々に促し表現から理解表現へと切り替える，という流れがあることが示唆される．

これらの結果は，実験としてはまだ不十分であり，より詳細な検証が必要ではある．一方でメッセージ入力部において理解表現を示す上では，理解の相槌を促しの相槌とともに用いることは，図 3 のように，徐々に「促しモード」から「理解モード」へと移行するような働きかけを行うことに対応する．つまり，このような相槌の使い分けは，最後のクロージング部への緩やかな接続を実現する上で妥当であると考えられる．そこで，4 章の実験におけるメッセージ入力部では，本節での扱ってきた

促しの相槌と理解の相槌を使い分けることとする．

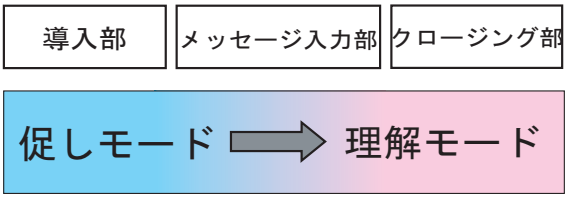

図 3　留守番電話での対話フェーズと対話姿勢との関係

4　対話実験による印象評価

本章では 3 章までの議論を踏まえて，「促しモード」から「理解モード」という流れに基づいた留守番電話の応答の有用性について検証を行う．ただし，3 章での留守番電話の応答に関して，導入部，メッセージ入力部における相槌，クロージング部のそれぞれの有効性について，全ての組み合わせを実験するのは被験者に負担がかかるため，被験者（話し手）の対話姿勢が切り替わると予想される部分に焦点を当てて対話実験を行うこととする．すなわち，導入部の有無により Disconnected モード→ Connected モードという切り替わりが生じるか，メッセージ入力部の相槌の有無により Connected モード→ Accepted モードという切り替わりが生じるかについて検証する．

4.1　実験内容

今回の実験では，被験者には導入部が有る場合，無い場合の 2 パターン，メッセージ入力部の相槌が無い場合，促しと理解の相槌を使い分けた場合の 2 パターンを組み合わせた 2 要因の被験者内計画で実験を行った．なお実験の詳細は以下のように設定した．

システムの音声：　電話対応に慣れた女性の録音音声，メッセージ入力部の相槌に関しては，促しの相槌は「はい」という音声を，理解の相槌は「あはい」という音声を用いた．

発声のタイミング：　導入部の音声に関してはユーザの発話終了から 1000msec のタイミングで挿入した．また，メッセージ入力部の相槌，クロージング部の音声は実験者が横で被験者の音声をモニターしながらキー操作により挿入した．

録音するメッセージ：　研究室見学についての詳細を聞くというシナリオで，最低限残してもらう内容を箇条書きで提示した．

被験者の印象評価：　それぞれの応答パターン（2 × 2 の 4 パターンの応答）に対して同じメッセージで録音してもらい，メッセージの残しやすさ，対話中にリラックスできたか，聞いてくれていると感じれるか，の 3 項目の印象評価を行った．評価の指標は 1 点から 7 点の 7 段階で，4 点を基準として，上に 3 段階，下に 3 段階として評価した．また，提示する順序によって慣れが生じ，評価結果に影響が出ると考えられるので，被験者ごとにラテン方格法により順序を入れ替えた．

以上の設定で被験者 16 名による対話実験を行った．

4.2　実験結果

表 1 に各評価項目の平均値，標準偏差を示す．各評価項目における主効果を検証するために，分散分析を行った．以下に各項目に関する要因の有意差を示す．

- メッセージの残しやすさ：導入部の有無（有意水準 3 %以下）
- リラックスできたか：導入部の有無（有意水準 1 %以下）
- 聞いてくれている感：導入部の有無（有意水準 1 %以下），相槌の有無（有意水準 1 %以下），導入部と相槌の交互作用（有意水準 3 %以下）

4.3　対話実験の考察

被験者の印象評価の平均値は全項目において，導入部と相槌を組み合わせることが最も良い評価となっている．導入部の影響が全体として評価に影響していることが窺える．特に，聞いてくれている感で有意差が出ていることから，Disconnected モード→ Connected モードの役割を導入部が果たしていると考えられる．さらに，メッセージの残しやすさという項目と，リラックスできたかという項目で有意差が出ていることから，Connected モードでの発話が，ユーザの負担を軽減していることもわかる．相槌に関しては聞いてくれている感において有意差が出ているが，その他の項目では，相槌がそれ程高く評価されていない．このことからユーザが発話をする上で必ずしも相槌が好ましいものであるとは言いがたい．それでも，導入部と相槌を組み合わせることで，全ての項目で最も高い評価となっていることから，導入部には相槌が雑音となる要素を軽減する効果があることもわかる．本研究においては，相槌に関して，促しの機能を持つものと理解の機能を持つものを切り替える方法で挿入を行ったが，促しのみの場合や理解のみの場合との比較もする必要がある．その点については今後追実験を行う予定である．

表 1　各評価項目ごとの平均値と標準偏差

		通常の留守番電話	相槌のみ	導入部のみ	導入部と相槌
メッセージの	average	4.56	4.5	5.25	5.38
残しやすさ	S.D.	1.73	1.46	1.39	1.49
リラックス	average	3.56	4.13	4.88	5.19
できたか	S.D.	1.69	1.73	1.58	1.47
聞いて	average	2.88	5.19	4.63	5.81
くれている感	S.D.	1.45	1.15	1.20	1.29

5　結論

本研究では，話し手の情報をうまく引き出す “聞き上手な対話システム” の設計を行った．その際に，聞き手が，話し手に対して情報取得への準備ができていると表現する「促しモード」から，話し手に対する理解表現を行う「理解モード」へとモードを切り替えることが，聞き上手につながるという仮設を提案した．具体的なアプローチとして，提案した仮説に基づいて，留守番電話録音状況でのシステムの応答を設計し，対話実験により有効性を検証した．その結果，メッセージの残しやすさ，リラックスできた度合，聞いてくれている感の 3 項目で提案したシステムが良い評価となることが確認された．また，分散分析の結果から，促しモードの機能を持つ導入部が全ての項目で有意な結果となっていた．このことから，導入部には，聞き手の存在を感じさせ，話し手の心的負担を軽減させる役割があることが確認できた．また，相槌に関しては，被験者全般に好ましい結果を与える訳ではなかったが，導入部と相槌の組み合わせが全項目で良い評価であったことから，促しモードから理解モードへの流れの有効性が検証できた．今回行った対話実験では，被験者にメッセージを最後まで残すというタスクで実験を行ったが，実際にシステムがユーザの発話を引き出すかを検証するためには，どの程度ユーザが本システムを用いて用件を残すかを検証する必要がある．そのために今後の課題として，タスクが課せられていないユーザが本システムでどの程度メッセージを残すかを検証する必要がある．

謝辞

本研究一部は科学研究補助金 18049046 の補助を受けて行った．

参考文献

[1] 松尾太加志: コミュニケーションの心理学, ナカニシヤ出版, (1999)

[2] 小島敬，川嶋宏彰，松山隆司: 情報爆発話時代におけるヒューマンコミュニケーション–聞き上手な対話システムの実現に向けて–, 情報処理学会第 70 回大会発表論文集, pp. 5_265–5_266(2ZL_1) (2007)

[3] 小林哲則, 藤江真也: マルチモーダル会話ロボット：ロボットが会話において行う「聴く」行為について, 計測自動制御学会誌, Vol. 46, No. 6, pp. 466–471, (2007)

[4] 竹内真士，北岡教英，中川聖一: 韻律・表層的言語情報を発話タイミング制御に用いた雑談対話システム, 情報処理学会研究会報告. SLP, 音声言語情報処理, Vol. 2004, No. 15, pp. 87–92 (2004)

[5] 向後千春，山西潤一: あいづち留守番電話の試作，日本認知科学会第 8 回大会発表論文集，pp. 72–73 (1991)

[6] 泉子・K・メイナード: 『会話分析』, くろしお出版，pp. 152–166 (1993)

[7] 新情報処理開発機構 (RWCP) 音声情報処理グループ, (1996)

[8] Weka Machine Learning Project, WEKA; http://www.cs.waikato.ac.nz/ml/weka

[9] 高梨克也，内元清貴，丸山岳彦: 『日本語話し言葉コーパス』における節境界認定, 平成 15 年度国立国語研究所公開研究発表会予稿集，pp. 35–36 (2003)