

話者の文脈とプロソディからの応答選択・調節の試み

塚原 渉¹
東京大学工学部

Abstract: 人間らしく応答する対話システムを構築するため、話者の心理状態(知識に関する自信の度合など)を推定し、それに応じて応答を選択することを目指している。今回、答えと答えの確認の繰り返しを基本的な会話パターンとするクイズ形式のタスクでコーパスを作成した。コーパスには、正解の場合の確認発話だけでも、「はい」「そう」「うん」、「答の繰り返し」「何も言わない」など、さまざまなものがあった。そこで、韻律情報と過去の発話履歴情報(過去の間違ひの回数、ヒントの回数)を用いて、知識に関する話者の自信の度合などを推定し、それに対応した応答を選択するための判断基準を提案した。

1 はじめに

音声対話システムを使いやすくするためには、協調的応答が重要である。多くの場合、協調的応答とは、ユーザの要求を満足させることと考えられているが、その他にも、応答そのものとして心地よい応答をするという点も考えられるであろう。前者はより正確にユーザの要求を把握することが解決課題となるが、後者は、例えば迅速な応答、あいづちを打つこと、いつも同じ答え方にならないようにすること、ユーザがより会話し易いような応答をすることなどが考えられる。

いままで、後者に関しては、言葉使いを変える(丁寧/ぞんざい)ことで印象が変化する(八木ら, 1995)、男声でなく女声を用いて応答するなどの工夫はある。しかし、人間対人間の会話においては、あいづちのような単純な応答に限って見ても、話し手の心理状態に応じた反応や、話し手の話に対する自分の態度を表すものなど、多様な応答が観察される(Ward, 1998)。それらの中には話し手を応援する態度を示し、話し手の発話を促すのに役立っていると思われるあいづちも多い。そこで、計算機が同じようなあいづちを打つために必要な情報の種類を明らかにすることが、本研究の目的である。

2 会話

分析を始めるに当たって、会話に次のような要素が含まれているものを対象とすることにした。

- 会話内容が単純である：最終的には、聞き手の役割をする計算機システムを作成する予定なので、音声認識が負担とならないためには単純な会話望ましい。
- 応答によって聞き手は積極的に話し手の発話を助けることができる：聞き手のあいづちが話し

手の状況に応じて変化し、また聞き手を助ける働きをするような会話望ましい。

これらを考慮すると、まず音声対話システムが仮定されるようなタスク(チケット予約など)では、会話内容の把握が本質的であるので、音声認識、対話管理の要素の方がより重要となり、不向きである。次に、マップタスクコーパスのような協調問題解決型会話では、話し手・聞き手の状態に応じたあいづちの応答が観察できるのだが、タスク自体の複雑さが大きいので、会話分析も計算機処理も難しくなる。自由対話も同様である。

一方で、家庭教師が生徒に簡単な問題を答えさせていく会話のようなケースであれば、生徒(話し手)の発話を促すようなあいづちの応答が見られ、かつ会話自体は比較的単純になる。そこで、このような場合を単純化した会話として、「山手線クイズ」を取り上げ、この会話を分析することにした。これは、「一人が山手線の駅名を順番に挙げ、もう一人がそれに対して正解・不正解の確認応答(「はい」「いや、ちがう」)をしたり、ヒントを出したりして、回答を助ける」というものである。会話例を示す。

【例1】

話し手：渋谷 えーとー 五反田 あ恵比寿
聞き手： はい ぶー はい

話し手：えーと恵比寿 五反田？ え？ 恵比寿？
聞き手： ぶー

話し手：恵比寿の次？ 何がある？
聞き手：恵比寿の次は？ ほら 目白じゃ

話し手： あ目黒か
聞き手：なくて はいはい

この会話をまず、41 対話(146 分) 収録した。それらの中から、話し手を積極的に助けて、協調的に会話しているかどうかという主観的基準で選んだ一人の聞き手について、この人物のあいづち応答を分析することとし、さらに

¹Wataru Tsukahara (tsuka@sanpo.t.u-tokyo.ac.jp)
Mech-Info Engineering, U. Tokyo, Bunkyo-ku Tokyo 113-8656
<http://www.sanpo.t.u-tokyo.ac.jp/~tsuka/>

	話し手の性別	話し手の立場	会話時間(分)	全回答数	平均回答時間(秒)
A	男	後輩	4.2	44	6.0
B	男	後輩	2.0	37	2.5
C	男	後輩	11.1	35	20.3
D	男	後輩	8.6	43	17.0
E	女	先輩	1.3	30	1.5
F	女	同期	2.8	40	4.5
平均	-	-	5.0	38.2	8.5

表 1: 分析した 6 会話の特徴

対話を収録した。表 2 に収録した会話の諸条件を示す。話し手(回答者)と聞き手(あいづち、ヒントを出す)の関係は、いずれも大学院の同じ研究室の知合いであった。聞き手、話し手いずれも音声対話の研究には関わっていない。

以下の分析はこれら同一の聞き手による 6 対話についてのものである。

3 あいづち選択の戦略

主なあいづちの種類を表 3 に示す。このなかから、主

応答	合計	A	B	C	E	F	G
駅名	16 (9%)			12	3		1
はい	82 (46%)	11	16	4	8	25	18
はい、正解	2					1	1
ほい	1						1
そー	10 (6%)	2	1	2	2		3
そーだ	2				1	1	
うん	19 (11%)	4	1	5	9		
はいはい	5	5					
そそそ	6	5					1
うんうん	1	1					
びんぼーん	10 (6%)		3	4	3		
いえー	1	1					
あー	2		1				1
おー	1						1
脇道	3	1	1		1		
無反応	9 (5%)		7		1		1
その他	6	3	1	1		1	1
全体	177	33	31	28	28	28	29

表 2: 会話に現れたあいづち応答の種類と出現頻度

要なあいづちとして、「はい」「そう」「うん」「(駅名の繰り返し)」「(なにも言わない)」を選んで分析し、計算機上であいづち生成を模擬するためのルールを作った。

3.1 ルールの作成方法

ルールを作るに当たって、話し手の回答²の韻律および、文脈情報を考慮した。具体的には、コーパスにおいて、

- 韻律情報： 回答のピッチ、パワー (全体および語末の高さ、分散、傾き) およびポーズ長、回答時間などの韻律情報
- 文脈情報： ヒント数、間違いの数、駅名の難易度³

を計算し、「はい」「うん」「(駅名)」を決定するルール(決定木)を予備的なルールとして作成した(表 3)。ルー

あいづち	条件
はい	一つ前も「はい」、間違い無し
(駅名)	長い回答時間
はい	「えーと」などの後、駅名が発話されている、難易度が高いのに少ないヒントで正解した
うん	「えーと」などの後、駅名が発話されている、難易度が高くてヒントも多く必要だった
はい	その他

表 3: あいづち選択の予備段階でのルール

ルは上から順に適用される。次に、このルールであいづち生成を模擬して合成した対話(対話 A を模擬)を 3 人の学生に聞いてもらい、不自然に感じる場所を指摘してもらった。そしてこれらの指摘を基にしてルールの修正を行なった。

3.2 あいづちを打たない

「連続して答えている場合にはあいづちは要らない」という指摘があった。良く駅名を知っている話し手の場合、早いペースで次々と駅名を言っていく。その際、聞き手はあいづちを打たずに、黙って聞いていることがコーパスからも観察された(6 例)。そこで、「あいづちを打つ瞬間⁴に話し手がすでに次の答を言い始めていたら、あいづちを打たない」ということにした。

【例 2】

話し手： 御徒町秋葉原神田東京

聞き手： おかちまち、あきはばら、
かんだ、とーきよー

3.3 元気な答えには元気なあいづちを打つ

「元気な答えには元気なあいづちを打つほうがよい」という指摘があった。主観的には、駅名の発話部分のピッ

² 正確には回答の駅名部分。【例】「えーと恵比寿か」→「恵比寿」

³ 7 人での聞きとり調査の平均、ほぼ一貫性があった。

⁴ あいづちのタイミングに関する予備実験の結果、回答部分の終了から 1 秒以内にあいづちを打てば、不自然に感じないことが分かったので、あいづちを打つ瞬間は、話し手の回答部分の終了の瞬間とした。

ちまたはパワーが平均的に高い場合元気な印象を感じることが分かったので、コーパスのデータについて聴覚印象に一致するように最適化した結果、話者の会話全体のピッチ、パワーのモードで正規化した

$$(\text{元気さ}) \equiv \bar{f}_{0\text{norm}} + 1.5\bar{E}_{\text{norm}}$$

が、3.5 より大きいものを元気な発話とした。この基準で測った元気な回答 (15 例) に対するコーパス中のあいづちを表 4 に示す。表 4 では、元気な答え (15 例) に対し

順位	元気さ	あいづち			
1	4.3	soso	9	3.6	pinpo-n
2	4.1	so	10	3.6	ununun
3	3.9	hai	11	3.5	so
4	3.9	ye-	12	3.5	so-da
5	3.9	pinpo-n	13	3.5	(無視)
6	3.8	pinpo-n	14	3.5	hai
7	3.7	haihai	15	3.5	sososo
8	3.7	sososo			

表 4: 元気な答えとあいづち

では、「うんうん」、「はいはい」、「そーそー」、「ピンポーン」「そーだ」「イエー」の計 10 例で、2/3 を占めていることが分かる。これらは、韻律的には元気なものではなかった⁵。むしろ語彙が普通の場合より変化していると考えられる。

そこで、普通のあいづち「はい」「そう」「うん」の代わりに、それぞれ「はいはい」「そうそう」、「うんうん」を「元気なあいづち」とし、これらを用いることにした。

3.4 あいづちを変えない

「順調に進んでいるのに、突然あいづちの種類が変わると不自然に聞こえる」という指摘があった。

【例 3】

話し手 : 新橋 有楽町 東京
聞き手 : うん うん ×はい
 ○うん

そこで、順調に進んでいる場合にはあいづちを変えないことにした。「順調」は、「前の正解のあいづちから 1 秒以内に答えていて、間違えていなく、ヒントも出されていない」場合とした。コーパスでは、この場合、前の応答が「はい」だった場合 (44 例) はそのうちの 40 例が「はい」になっていた。「はい」が順調な場合に多く用いられることが分かる。他に、「うん」は 2 例中 1 例が、「(駅名繰り返し)」は 1 例中 1 例がこの場合に当たっていた。

⁵元気さとの相関は無かった

3.5 難しい場合には優しく

次の駅名が難しくて分からなかった場合、回答までに時間がかかる。回答までの時間 (filled/unfilled pause) が長いほど、話し手は答えに関して良く分かっている (Brennan and Williams, 1995)。従って、優しい応答が望ましい。コーパスでは、回答までの時間が長い場合には、沈黙の他には filler 語、ヒント、関係のない話題などがあり、この場合、あいづちとして、駅名を繰り返していることが多かった (例: 括弧内は休止区間の秒数)。

【例 4】

話し手: (7s) (1.5s) 五反田
聞き手: 数字で言うと 5 五反田

回答時間が 30 秒を越える場合 (全 15 例) には、駅名の繰り返しが多い。「ピンポーン」というあいづちが多い。「そー(だ)」が 2 回だった。そこで、前のあいづちからの次の回答までの時間が非常に長い場合 (30 秒以上) には、駅名を繰り返すこととした。

このあいづちは、単なる「はい」などよりも優しく、「そう」が褒める要素が強い (後述) のに比べると、回答者を急かさず、同じペースで進むという印象を与える⁶。

3.6 良くできた場合は褒める

答が順調に出てこない場合には、ヒントが出されることが多い。「ヒントがうまく正解につながった際には、順調な場合と同じあいづちを出すのではなく、『良く出来た』と褒めてあげるようなあいづちが望ましい」との指摘があった。そこで、ヒントを出された後に正解できた場合には、「そう」と答えることにした。コーパスでは、ヒントを一回以上出された場合のあいづちは「うん」(9 回)、「(駅名繰り返し)」(9 回)「ピンポーン」(7 回)「そう」(5 回)「はい」(5 回)、「そうそう」(4 回)の順であった。したがって、必ず「そう」が選ばれるとは言えない。しかし「そうそう」も含めると、「そう」「そうそう」のうちの半数以上がこの場合に含まれていた。会話の例を示す。

【例 5】

話し手 : 西日暮里
聞き手 : ほら、日暮里に近いんだよ そう

後述する「うん」は、優しい印象を与えるが、「そう」や「そうそう」は、むしろ聞き手の喜びを伝える印象があった。

3.7 自信のない口調の場合には優しく

「順調な場合と同じあいづちでは、冷たい印象を与える」と指摘された部分があった。これらは、ヒントなしで正解した場合に、自信がなさそうな口調で答えた場合だった。聴覚印象では、これらは「疑問のイントネーション」の場合であった。

⁶これに対して「うん」にはこのような積極的な意味づけは感じられなかった

⁷本コーパスでは、疑問のイントネーションは、「発話の駅名部分の最後の 1 分の 1 部分のピッチの傾きが大きい」と単純化できる。

【例6】

話し手：日暮里？

聞き手： うん

そこで、自信のない答えが正解だった場合には、より好ましいと指摘された「うん」を用いることにした。コーパスにおいては、自信のない口調の回答(全20例)に対するあいづちの種類は、はい(12)、うん⁸(4)、はいはい(3)、そう(1)などであった。

話し手(回答者)はあまり良く分かっていない点では3.5と同じだが、3.5の場合はヒントが与えられることが多いので、回答する際にはある程度答えに自信があると考えられる。しかし、この場合には回答した段階ではまだ自信は無いという点が異なり、その結果異なるあいづちが用いられたと考えられる。

3.8 デフォルトの「はい」

通常会話の最初の方などでは、以上のどの場合にも当てはまらないことがある。そこで上のどれにも当てはまらない場合には、デフォルトとして、もっとも普通の確認のあいづちである「はい」を用いることにした。

3.9 ルールのまとめ

以上をまとめたものを表5に示す。 結果として、

章	あいづち	条件	回答の特徴	あいづちの特徴
3.2	(なし)	早いペース	極めて順調	
3.3	はいはい そうそう うんうん	ビッチ、パワー	元気	元気に
3.4	(前と同じ)	短い間、間違い無し、ヒント無し	順調	
3.5	(駅名)	長い回答時間	難しい	優しく
3.6	そう	ヒント後		褒め、喜ぶ
3.7	うん	疑問調、ヒント無し	自信が無い	優しく
3.8	はい	その他		デフォルト、事務的

表5: あいづち選択のルール

3.1の予備的ルールと比較してみると、3.5は予備実験の段階から変わらず、3.4は元々「はい」のみを対象としていたのが拡張された。3.6、3.7は新たに加わったが、コーパスから導いた予備的ルールの単純化と解釈できる。3.2、3.3は全く新たに加わったものである。

このルールで排他的に競合するのは、「(駅名)」と、「そう」または「うん」であるが、その場合には「(駅名)」を暫定的に用いることとした。

⁸ haihaiun も含む

分析対象としたコーパスについての、このルールに基づいた予測の結果を表6に示す。 予備的ルールに比

	はい	そう	うん	駅名	そうそう ⁹	無視	合計
	82	10	19	16	15	9	177
予備	80 (98%)	0 (0%)	2 (10%)	6 (38%)	0 (0%)	0 (0%)	88 (50%)
今回	51 (66%)	3 (30%)	4 (21%)	7 (43%)	5 (38%)	6 (66%)	79 (45%)

表6: 予測の結果(6対話)

べ、「はい」の予測率が下がったが、それ以外のあいづちが予測できるようになった。

4 まとめ

話者の韻律情報や過去の発話履歴から、話者の自信の度を推定し、それに応じて、話者をより応援するようなあいづち応答を行なうための、応答選択のルールを提案した。

今後は、これらのルールを用いて合成した対話を複数人に聞いてもらい、それぞれのあいづちに不自然さが感じられないかどうかを確認する予定である。さらに、実時間で聞き手の役割を模擬するシステムを作成することを考えている。そして、被験者に実際にクイズを行なってもらい、あいづちを変化させることによって、どの程度話し手を応援するシステムとなり得るかを検討したい。

対象とした会話は山手線のクイズという特殊な会話形式だが、聞き手があいづちによって積極的に話し手を応援する場面は一般の会話でもあり得るので、そのような場合にはこれらの応答戦略が利用できるのではないかと期待している。

【謝辞】

本研究を行なうにあたり、財団法人中山隼雄科学技術文化財団の援助を受けた。記して謝意を表する。また、指導教官として日頃御指導を頂く東京大学のNigel Ward 助教授に深く感謝致します。

【参考文献】

- S.E. Brennan and M. Williams. 1995. "The feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive Status of Speakers". *Journal of Memory and Language* 34. pp. 383 - 398
- Ward, Nigel. 1996. "The Relationship between Sound and Meaning in Japanese Back-channel Grunts". 言語処理学会第4回年次大会発表論文集(採録予定)
- 八木正紀、平栗覚、伊賀総一郎、安村通晃.1995. "音声対話におけるエージェントと人間の発話の関連". 情報処理学会研究報告 SLP-7-14. pp. 85 - 90

⁹ はいはい、うんうん、そうそう、そーだ、そー(駅名)を含む