

話しことば対話における協同理解戦略

堂坂 浩二 川端 豪

NTT 研究所

1 はじめに

話しことば対話は、人が他者とともに行う協同活動である。協同活動は個々の協同行為の実施を積み重ねていくことにより展開していく。協同行為とは、複数の主体が、互いの個別行為の間での調整を計りながら、全体で一緒になって実施する行為である[1]。

対話においては、様々な協同行為が実施される。本稿では、一方の対話参加者が行なった発話の内容と意図に関して対話参加者の間で合意することを目的として実施される協同行為に注目する。この協同行為を協同理解と呼ぶ。たとえば、一方の参加者 A が発話を行ない、他方の参加者 B が、それに対して、間投詞（例：「はい」）や A の発話の一部を繰り返すことによる応答を行ない、さらに A が B の応答を評価して、必要ならば B の応答に対して応答するといった過程を全体で一つの協同行為と捉えて、協同理解と呼ぶ。協同理解のモデルは、音声対話システムがユーザと協同でタスクを達成する能力を備えるために必要不可欠である。

そこで、協同理解のモデルを構築するための予備的考察として、話しことば対話コーパスにおいて、人が採る協同理解戦略について分析を行った。対話の状況を分類し、各状況においてタスクの達成のために必要な情報を一方が送信し他方が受信するときに、情報の送信者と受信者のそれぞれが採る戦略を分析した。対話の状況に応じて、情報の送信者と受信者の双方が協同理解のために費やす労力は変わることを示す。

2. 関連研究

協同行為をモデル化するための具体的な理論としては、協同行為に従事するための行動計画を記述、計算するための理論[3]がある。また、ゲーム理論は、複数の主体が協力、競合する状況をモデル化するための理論である。実際、ゲーム理論を使って、自然言語によるコミュニケーションの図式をモデル化することも試みられている[4],[7]。本稿では、厳密な理論に基づいて現象を分析することはしない。むしろ、対話コーパスの分析に基づいて、現実に行う協同理解戦略の実際を観察して知見を得ることが目的である。

話しことば対話では、書きことばにおける文よりも小さな文法単位が発話単位として使われる[2]。そういった小さな発話単位を使って、発話は漸次的に生成される。漸次的な生成は、話し手が協同理解の過程に従事するには有利な戦略である。なぜなら、漸次的に発話を生成することにより、相手の理解状態を確認しながら情報を伝達できるからである。しかし、漸次的に発話を生成するというだけでは、片側の対話参加者のみの振る舞いを規定しているだけであって、二人の対話参加者が一緒になって協同行為を実施するための戦略としては不十分である。協同理解戦略を分析するためには、2 人の対話参加者が、ある対話状況において、どういった戦略を採り合うかということの問題にしなければならない。状況に応じて人が振る舞いを変えるのは自然なことだから、一方が漸次的に発話を生成する際の他方の振る舞いは、対話状況によって異なることが予想されるし、また、対話状況によっては、同じ漸次的な生成といっても異なる方略が採られる可能性がある。本稿では、このことをコーパス分析により確認する。

音声対話システム DUG-1[6]は、ユーザ発話の逐次的理解[5]とシステム発話の漸次的生成[2]によって、ユーザ発話を理解しながら間投的発話（例：「はい」）を行なうことや、システム発話途中でのユーザ発話に応じて話の進め方を調節することができる。このことにより、DUG-1 は、特定の対話状況においては、ユーザと一緒に協同理解に従事することができる。しかし、DUG-1 は、あらゆる対話状況について、状況に応じた協同理解戦略を備えているわけではないし、既存の対話システムで状況に応じた協同理解戦略を備えたものはない。対話状況に応じた適切な協同理解戦略を備えた音声対話システムを開発するための第一歩として、本稿では、人が状況に応じて異なる協同理解戦略を採ることを示す。

3. 対話コーパス

分析に用いた話しことば対話コーパスは、TV 番組録画予約コーパスと経路案内コーパスである。

TV 番組録画予約コーパスは次の対話実験により収集した。実験には 34 人の被験者が参加し、136 セッションの対話が収録された。各対話において、二人の被験者(U と S)が非対面で会話をを行った。対話の

中で、UがSに対して、TV番組の録画予約を依頼することが要請された。Sに前もって与えられた情報は各番組の知識である。番組の知識としては、番組のタイトル、サブタイトル、ジャンル、出演者、放送時間がある。一方、Uに前もって与えられたのは、録画すべき番組の知識の一部である。特に、Uは番組のタイトルは知らされていない。したがって、Uが自分の知っている番組の知識を一方的にSに伝えるだけでは、録画すべき番組を一意に同定できるとは限らない。SとUは互いのもっている情報をやり取りすることにより録画すべき番組を決定する。

経路案内コーパスは、次の対話実験により収集した。実験には55人の被験者が参加し、51セッションの対話が収録された。各対話において、二人の被験者(UとS)が非対面で会話をを行った。対話の中で、SがUに対して、ある場所から別の場所へ行くための経路を教示することが要請された。Sは経路を見つけるために十分な知識をもっており、Uはもっていないように選ばれた。

分析のために、それぞれのコーパスから5対話を選んだ。便宜上、Sをシステム、Uをユーザと呼ぶ。

4. 分析対象と分析方法

本稿の目的は、対話状況に応じて協同理解戦略が変わることを示すことにある。次の対話状況を取り上げる。

- [TV-ユーザ] TV番組録画予約において、ユーザが録画したい番組の知識をシステムに伝えている対話状況
- [TV-システム] TV番組録画予約において、システムがユーザの希望に沿う番組の知識をユーザに伝達している対話状況
- [経路-システム] 経路案内において、システムがユーザの希望に沿う路線経路の知識をユーザに伝達している対話状況

ここでは、タスクを達成するために必要な情報(タスク情報)のやり取りに関する協同理解の戦略を分析する。様々な種類のタスク情報がある。ここでは、伝達されたタスク情報の量をできるだけ客観的に見積もるために、次のタスク情報を分析対象とした。TV番組録画予約においては、番組のタイトル、ジャンル、出演者といった固有名、経路案内においては、路線、駅、バスの名前といった固有名をタスク情報とした。タスク情報を表す名詞1個につき1個のタスク情報と数える。

タスク情報の伝達は、タスク情報を1個以上含む発話単位ごとに行なわれる。発話単位は次のように定義した。

- 節は発話単位である。
- つなぎ語は発話単位の区切りである。
- 400msec以上のポーズは発話単位の区切りである。

さらに、旧情報を担う名詞句によって表されたタスク情報は、タスク情報を伝達するためではなく、文脈の首尾一貫性を維持するために使われている可能性があるので、分析対象から外した。旧情報を担う名詞句とは、代名詞、指示的な名詞句、既出対象を参照する提題化された名詞句である。

ここで分析する協同理解の過程は、ある対話状況でタスク情報の送信者と受信者がタスク情報のやり取りに関して協同理解に従事しているという過程である。ここで取り上げた各対話状況では、送信者と受信者は固定されている。対話状況[TV-ユーザ]では、送信者はユーザ、受信者はシステムであり、対話状況[TV-システム]、[経路-システム]では、送信者はシステム、受信者はユーザである。送信者はタスク情報を含む発話単位を発する。受信者は、それに対して、タスク情報を受け取ったことを示す応答を行なう機会をもつ。この応答を承認応答と呼ぶ。承認応答は、「はい」といった間投詞や、タスク情報を表す名詞の繰り返しによって行なわれる。

ここで用いた分析方法では、承認応答を数えることが必要になるので、承認応答の数え方を次の対話を例にとって説明する。

- (1)S:「名作劇場」という番組と「アンコール映画」という2つの番組があります。

(2)U: はい

(3)S: 最初の方は、午後十時からの「名作劇場」で

(4)U: 「名作劇場」ですね

これは、対話状況[TV-システム]における情報のやり取りである。発話単位(1)はタスク情報の送信者Sの発話であり、「名作劇場」と「アンコール映画」という2個のタスク情報(番組タイトル)が現れる。発話単位(1)に対して、受信者Uは間投詞「はい」による承認応答(2)を行なっている。発話単位(1)における2個のタスク情報の出現に対して、それぞれ1/2個の承認応答があったと数える。発話単位(3)では、「午後十時」(放送時間)と「名作劇場」(番組タイトル)の2個のタスク情報が現れる。これに対して発話単位(4)で名詞の繰り返しによる承認応答が行なわれている。発話単位(3)におけるタスク情報「名作劇場」の出現に対する承認応答の数は1個である。タスク情報「午後十時」の出現に対する承認応答の数は0個である。また、タスク情報「名作劇場」に対して対話状況を通して成された承認応答の総数は3/2個(1/2+1)である。

分析方法を説明する。タスク情報のやり取りに関して協同理解が実際に成功したかどうかという観点から分析することは難しい。そこで、ここでは、タスク情報の送信者と受信者が協同理解に費やす労力

の大きさを分析した。次のことを調べた。

- ・（分析1）送信者の発話単位に含まれるタスク情報の数
- ・（分析2）送信者の発話単位に含まれるタスク情報の個々の出現に対して成された受信者からの承認応答の数
- ・（分析3）同一のタスク情報に対して対話状況を通して受信者から成された承認応答の総数

（分析1）では、受信者からの承認応答は送信者の発話単位ごとに行なわれるので、送信者の発話単位に含まれるタスク情報の数が少ないほど、送信者はタスク情報1個あたりの協同理解に費やす労力を大きく見積もっていると考えられる。

（分析2）では、送信者の発話におけるタスク情報の1回の出現に対する受信者からの承認応答の頻度が高いほど、それだけ送信者と受信者が協同理解に費した労力は大きいと考えられる。ここで、タスク情報の個々の出現に対する承認応答の数が1個である場合が全体に占める割合を協同理解の有効度とよぶ。協同理解の有効度が高いほど、タスク情報の1回のやり取りに関して送信者と受信者が協同理解に費した労力は高い。

（分析3）は、送信者と受信者が対話状況を通して協同理解に費した労力を調べるために行なう。ここで、対話状況を通して同一のタスク情報に対する承認応答の総数が1以上になった場合が全体に占める割合を、協同理解の累積有効度と呼ぶ。協同理解の累積有効度が高いほど、送信者と受信者の双方が対話状況を通して協同理解に費した労力は大きい。特に累積有効度がある有効度よりも高くなるのは、送信者が対話状況を通して同一のタスク情報を何度も伝達し、受信者がそれに合わせて何度も承認応答を行なったことを示す。

5. 分析結果

（分析1）の結果を示す。対話状況[TV-ユーザ]において、送信者であるユーザの発話単位の総数は36個であった。そのうち、タスク情報1個を含む単位は31個、タスク情報2個を含む単位は5個であった。発話単位1個あたりのタスク情報数の平均値は1.1、標準偏差は0.35である。対話状況[TV-システム]において、送信者であるシステムの発話単位の総数は69個であった。そのうち、タスク情報1個を含む単位は46個、タスク情報2個を含む単位は18個、タスク情報3個以上を含む単位は5個であった。発話単位1個あたりのタスク情報数の平均値は1.5、標準偏差は0.84である。対話状況[経路-システム]において、送信者であるシステムの発話単位の総数は88個であった。そのうち、タスク情報1個を含む単位は71個、タスク情報2個を含む単位は17個であった。発話単位1個あたりのタスク情報数の平均値は

承認数	0	1/2	1	計
頻度	2	6	33	41

表1 対話状況[TV-ユーザ]において、タスク情報の個々の出現に対する承認応答の数の頻度分布

承認数	0	1/6	1/3	1/2	1	計
頻度	25	6	5	24	41	101

表2 対話状況[TV-システム]において、タスク情報の個々の出現に対する承認応答の数の頻度分布

承認数	0	1/2	1	計
頻度	16	18	71	105

表3 対話状況[経路-システム]において、タスク情報の個々の出現に対する承認応答の数の頻度分布

承認総数	0	1/6	1/3	1/2	1
頻度	18	6	5	18	35
承認総数	3/2	2	計		
頻度	2	3	87		

表4 対話状況[TV-システム]において、同一タスク情報に対する承認応答の総数の頻度分布

承認総数	0	1/2	1	3/2	2	5/2
頻度	1	5	43	3	4	5
承認総数	3	7/2	計			
頻度	2	1	64			

表5 対話状況[経路-システム]において、同一タスク情報に対する承認応答の総数の頻度分布

1.2、標準偏差は0.40である。

（分析2）の結果を示す。各対話状況において、タスク情報の個々の出現に対する承認応答の数の頻度分布を表1、表2、表3に示す。協同理解の有効度は、対話状況[TV-ユーザ]では0.80(33/41)、対話状況[TV-システム]では0.41(41/101)、対話状況[経路-システム]では0.68(71/105)である。

（分析3）の結果を示す。対話状況[TV-システム]と[経路-システム]のそれぞれにおいて、同一タスク情報に対する承認応答の総数の頻度分布を表4と表5に示す。対話状況[TV-ユーザ]では、同一タスク情報が繰り返し言及されることはほとんどなかった。すなわち、（分析2）と結果は変わらない。協同理解の累積有効度は、対話状況[TV-ユーザ]では、有効度と変わらず、0.80であり、対話状況[TV-システム]では0.46(40/87)、対話状況[経路-システム]では0.91(58/64)である。

6. 考察

(分析1)の結果を考察する。発話単位1個あたりのタスク情報の数が少ないほど、協同理解には有利であるから、タスク情報の送信者がタスク情報1個あたりの協同理解に費やすのに見積もっている労力は、対話状況[経路-システム]、[TV-ユーザ]の方が対話状況[TV-システム]よりも大きい。実際、対話状況[TV-システム]では、タスク情報6個を含む発話単位さへ現れ、他の2つの対話状況に比べて、協同理解に費やす労力は小さく見積もられている。

(分析2)の結果では、協同理解の有効度は、対話状況[経路-システム]、[TV-ユーザ]の方が対話状況[TV-システム]に比べて大きい。タスク情報のやり取りが1回行なわれるたびに、その協同理解に費やされる労力は、対話状況[経路-システム]、[TV-ユーザ]の方が、対話状況[TV-システム]に比べて大きい。

(分析3)の結果において、協同理解の累積有効度は、対話状況[経路-システム]、[TV-ユーザ]が、対話状況[TV-システム]に比べて大きい。対話状況を通して協同理解に費やされる労力は、対話状況[経路-システム]、[TV-ユーザ]の方が、対話状況[TV-システム]に比べて、大きいことがわかる。

(分析2)において、協同理解の有効度は対話状況[TV-ユーザ]の方が対話状況[経路-システム]よりも大きい。これは、TV番組録画予約タスクでは、経路案内タスクに比べて、送信者が対話状況を通して伝達すべきタスク情報の総数が比較的少ないために、同じタスク情報を何度も繰り返す必要性が小さいからであろう。一つの対話状況を通して伝達されるタスク情報の総数は、対話状況[TV-ユーザ]では平均7.8個、対話状況[経路-システム]では平均12.8個である。このことは、対話状況[経路-システム]では、協同理解の累積有効度が(分析2)の有効度よりかなり大きくなっているが、対話状況[TV-ユーザ]ではその差が見られないことから分かる。

対話状況[経路-システム]と[TV-ユーザ]は、タスクが異なっており、またタスク情報の送信者も異なっているにもかかわらず、協同理解に費やす労力は大きいという共通点があり、それに対して、対話状況[経路-システム]では、その労力が小さい。

この結果は、タスク情報1個あたりの価値が対話状況によって異なるためではないかと推測される。タスクを成功させるためにタスク情報が必要とされる度合いが高いほど、タスク情報の価値は高いとする。対話状況[経路-システム]、[TV-ユーザ]では、それぞれシステムが伝達する路線や駅の名前、ユーザが伝達する録画したい番組の知識は、対話参加者の間で伝達が成功したことを確認する必要性が高い。そうしなければタスクが簡単に失敗するからである。ゆえに、この2つの対話状況では、タスク情報1個あたりの価値が高い。これに対して、対話状況[TV-

システム]では、システムがユーザに伝達する番組の知識のすべてがタスクを成功させるために必要な情報とは限らない。ユーザは、システムが伝達する多くの番組情報からタスクを成功させるために必要な情報を選択すればよいだけである。この意味で対話状況[TV-システム]では、他の2つの状況に比べてタスク情報1個あたりの価値は小さい。

音声対話システムは、対話状況によって協同理解に費やす労力を調節する必要がある。さもないと、やり取りされる情報の価値に不釣り合いなほど、協同理解に費やす労力が大きすぎて冗長なやり取りが生じてしまったり、協同理解に費やす労力が小さすぎてタスクの達成が簡単に失敗してしまうだろう。

7. おわりに

本稿では、対話コーパスの分析に基づいて、タスクの達成のために必要な情報のやり取りに関する協同理解に費やされる労力は、対話状況に応じて変わることを示した。やり取りされる情報の価値が大きいほど、大きな労力を費やして協同理解が行なわれるとの推測を得た。

謝辞

日頃より有益な示唆を頂くメディア情報研究部萩田紀博部長、相川清明リーダー、マルチモーダル対話研究グループの諸氏に感謝致します。

参考文献

- [1] Clark H.H.: Using language. Cambridge University Press, 1996.
- [2] 堂坂浩二、島津明: タスク指向型対話における漸次的発話生成モデル. 情報処理学会論文誌 37(12): 2190-2220, 1996.
- [3] Grosz B.J., Kraus S.: Collaborative plans for complex group action. Artificial Intelligence 86: 269-357, 1996.
- [4] Hasida K.: Issues in Communication Game. Proceedings of the 16th International Conference on Computational Linguistics (COLING-96): 531-536, 1996.
- [5] 中野幹夫、宮崎昇、平沢純一、堂坂浩二、川端豪: 多重文脈を用いた逐次的な発話理解. 情報処理学会研究会報告 SLP-22: 21-26, 1998.
- [6] 中野幹夫、堂坂浩二、宮崎昇、平沢純一、田本真詞、川森雅仁、杉山聡、川端豪: 柔軟な話者交代を行なう音声対話システム DUG-1. 言語処理学会第5回年次大会, 1999.
- [7] Parikh P.: Communication and strategic inference. Linguistics and Philosophy 14: 473-514, 1991.