

# 対話行為と局所対話構造の認識

高野 健治 島津 明

北陸先端科学技術大学院大学 情報科学研究科

## 1 はじめに

自然言語での対話システムを構築するための必要な機能の一つに対話行為の認識がある。発話は話し手が聞き手に対して何らかの意図を持ってなされる行為である。話し手の対話行為を理解することは、対話を円滑に行うためには必須の機能である。例えば、「はい」という発話は状況によって質問に対する肯定的な返答行為に使われたり、単にあいづち行為として使われたりする。この行為の認識を間違えたならば、対話はおかしなものになる。

対話行為の認識については、プランベースによる研究がある [1]。発話の表層情報から直接、対話行為を求める研究がある [2][3]。近年では、計算機の高性能化とコーパス環境の整備から、n-gram モデルを用いた表層情報からの認識方法が研究されている [4]。適切な認識や発話生成のためには部分対話などの談話構造の認識が必要である [5] が、n-gram モデルを用いた方法では部分対話を捉えていないことから、問題がある。対話構造を考慮した研究として [6] がある。これは対話の内容を木として捉えているが、発話タイプの種類が非常に少ないため、対話行為の認識としては問題がある。

我々は、対話行為の認識に、対話構造を考慮したモデルを用いるため、対話行為と対話構造との関係を主な対象として分析している [7]。その結果、交通経路案内対話では局所レベルで対話行為の系列に構造が存在し、この局所構造はほぼ交換 [8] 単位に相当することが分った。

本報告では、局所対話構造の認識と発話の対話行為を同時に認識することを目指し、対話行為と局所対話構造の関係の規則、認識方法などを提案し、実験結果を示す。

## 2 局所対話構造による対話行為認識モデル

対話行為認識システムは対話を漸次的に処理する必要がある事を考慮し、発話単位 [9] を処理単位として対話行為の候補を推測して、局所的な部分対話構造を認

7 D: この愛甲石田というのが、〈はい〉えっとー厚木、  
NTT 厚木研究所の最寄り駅になるんですよ。

8 A: はい。

↓

7-1 D: この愛甲石田というのが、

7-2 A: はい

7-3 D: えっとー

7-4 D: 厚木、

7-4 D: NTT 厚木研究所の最寄り駅になるんですよ。

8-1 A: はい。

図 1: 発話単位への分割例

識し、結果として対話行為を認識するモデルを考える。

### 2.1 発話単位

対話において話者は伝達すべき情報を一文にまとめて発話するのではなく、情報を複数の単位に分け、その単位毎に生成していることが観察される [9]。堂坂らの研究に従い、この単位を発話単位と呼び、我々のモデルでの発話処理単位とする。

発話は相手発話によって割り込まれることから、そのため元の発話は不完全な形で分割され、独立した発話単位と認めにくいものが生成される場合があるが、そのような発話も擬似発話単位として扱い、後の対話行為タグ付の際は専用のタグ“%”を付加している。例えば以下のように発話単位に分割された時に付加される。( ) 内が対話行為タグとする。

K: 慶応大学の古積です (情報伝達)

D: あっ (フィラー)

K: けど. (%)

図 1 は発話を発話単位に分割した例である。分割したものには、“元の発話番号 – 分割した順番”という形で ID が付けられる。

### 2.2 対話行為

発話単位毎に付加する対話行為タグは、荒木らの発話単位タグ標準化案 [10] をもとにし、実際にコーパス

- 11 K: えーっと、今どちらですか？  
 12 M: 今は自宅です。  
 13 K: えーと自宅と言うのはどちら [1]  
 14 M: [1] えーと三鷹です。  
 15 K: 三鷹ですか？  
 16 M: はい。

図 2: 交換構造と局所対話構造が一致しない例

を解析していった結果、不足していると思われるタグを加えた。具体的には、“フィラー”、“あいづち”、“%”である。これらは厳密には対話行為とは言えない。しかし、本研究では各発話単位毎に何かしらのタグを出力する事を想定していることと、対象となる発話が一見対話行為なのか否かがわからない発話があるために付加した。例えば、発話“はい”を考えた場合、この発話は Yes/No の質問に対する Yes の回答の“はい”なのか、または単にあいづちで“はい”と言っているだけなのかは、対話行為を正しく認識しない限りわからない。

## 2.3 局所対話構造

対話全体を一つの構造として捉えようとした場合、一般にあいまい性があるが、局所レベルに着目してみると、対話行為の列としてのパターンがみられることが多い。本研究では、話者が伝えようとする情報について、相手と相互信念の基盤化 [11] がなされたと認められるまでの発話列を局所対話構造とする。基盤化がなされたと認められるまでの発話の並びの多くは交換構造である。しかし、常に交換構造と一致するわけではない。図 2 の対話は、二つの交換構造、発話 11~12 と 13~16 からなる。発話 11~16 の対話は全体で話者 M の現在地の情報が基盤化されたとみられるので、一つの局所対話構造として扱う。

図 3 は図 1 の発話列に局所対話構造タグまで付けたものである。各発話単位の後に () で括弧であるものが対話行為で、冒頭の“1 → 2:情報伝達”が局所対話構造名である。この局所対話構造では話者 1 が話者 2 に対して情報伝達行為を行うことを目的とした部分の対話であることを示している。交通経路案内対話において話者 1 は経路の案内者側、話者 2 はその説明を受ける側としている。矢印の元が作用を及ぼす側(開始発話者。図では話者 D で、D は経路の案内者である)、矢印の先が作用を及ぼす対象で、その後この局所対話構造を代表する対話行為を記し、これがこの局所対話構造の働きとなる。多くの場合、構造の働きは開始発話者の最初の発話の対話行為になる。

1 → 2:情報伝達

7-1 D: この愛甲石田というのが、(情報伝達)

7-2 A: はい (あいづち)

7-3 D: えっとー (フィラー)

7-4 D: 厚木、(情報伝達)

7-5 D: NTT 厚木研究所の最寄り駅になるんですよ。(情報伝達)

8-1 A: はい。(肯定・受諾)

図 3: 局所対話構造タグの例

## 2.4 局所対話構造の規則化

局所対話構造までタグ付けしたコーパスを利用し、局所対話構造の規則化を行う。局所対話構造名を左辺、その構成要素としての対話行為の列を右辺として、コーパスから抜き出したものを式にする。(1) 式は図 3 から抜き出したものである。

{1 → 2:情報伝達} ⇒ 1:情報伝達 2:あいづち 1:情報伝達 1:情報伝達 2:肯定・受諾 …(1)

フィラー発話は規則に含まない。フィラーは対話行為ではないことと、音声情報を利用した際、その情報によりフィラー発話である事が認識できると考えるためである。

コーパス量の少なさから抜き出した規則だけでは必ずしも量が十分ではない。よって、規則の拡張を行う。具体的には以下に基づいた拡張化を行う。

- 本来一つの対話行為となる発話でも、いくつかに分割されて発話される。よって同じ対話行為の繰り返しを許容する。
- 開始発話は相手のあいづち発話によって分割されるため、発話に対して任意に相手のあいづち発話の挿入が可能にする。
- 補足発話(了解)は相手応答発話または補足発話に続くことができることから、補足発話が後に続くことを可能にする。

以上の拡張作業の後、基となった構造規則が受理する発話列は拡張規則で受理できるため、除去する。

## 2.5 局所対話構造を用いた対話行為認識

本研究での対話行為と局所対話構造の認識方法の主な流れは以下である。まず、発話の表層情報から候補となる対話行為を何らかの方法で求める。これにより求められた複数の対話行為候補の列に対して局所対話構造規則を用い、対話行為を絞り込み、局所対話構造を求める。

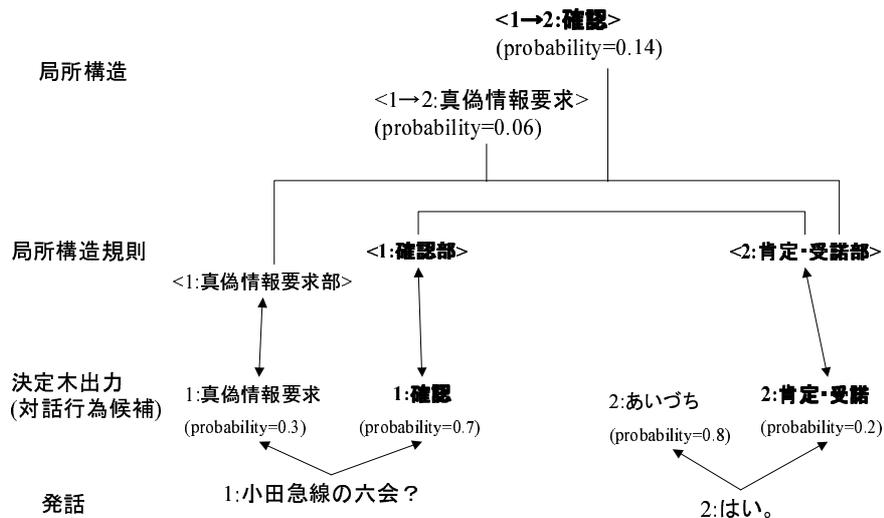


図 4: 認識方法イメージ

図 4 は本モデルでの認識方法を表す。まず、話者 1 の発話「小田急の六会？」という発話に対し、二つの発話の対話行為候補として「真偽情報要求」と「確認」が出力される。probability はその候補の確率を示す。同様に話者 2 の発話「はい」に対しては対話候補「あいづち」と「肯定・受諾」を出力する。以上の対話行為の組合せが可能な局所対話構造規則を当てはめた結果、ここでは「1 → 2: 確認」と「1 → 2: 真偽情報要求」の二つが該当したとすると、各対話行為の組合せの確率が高い方を正しいものとし、図 4 では最初の話者 1 の発話は「確認」、話者 2 の発話は「肯定・受諾」、局所対話構造は「1 → 2: 確認」と認識する。

### 3 実験と評価

提案モデルの評価のために以下の実験を行った。最初に、発話から対話行為候補を選択するための決定木を作るために、C4.5 アルゴリズムを用いた学習を行った。学習要素としては発話単位毎に最後の文節の形態素解析結果と話者情報を用いた。交通経路案内の計 9 対話 (1,776 発話単位) 分のタグ付きコーパスを学習元とした。

次に学習元のコーパスから抽出した局所対話構造の規則の拡張を行った。コーパスから抜き出した規則は 269 で、拡張した結果は 411 になった。

以上を利用し、3 対話に対して認識実験を行った。今回の実験では、規則の確率は求めず、対話をセグメントした上で規則を使用した。表 1 は認識結果 (Recall 値) を示す。“baseline”は、それぞれの話者の全発話を、学

習元の対話で最も多かった発話にしたものである。経路案内者側の全発話を「情報伝達」、説明を受ける側の全発話を「あいづち」にした。“決定木”は、C4.5 を用いた学習結果の決定木を用いて最も確率が高かったものである。“決定木 (直前対話行為利用)”は、洪 [12] にならぬ決定木学習の際に直前の対話行為情報も含んだものである。“局所対話構造”は“決定木”を用いた後、局所対話構造規則を適用したものである。“局所対話構造”のみ、局所対話構造が認識できなかった場合は出力がないことから、recall と precision に差が出るため、() 内に precision の値を記している。“複合型”は基本的に“局所対話構造”を用いるが、構造規則が適用できなかった場合は“決定木”の結果を用いる。

表 2 は局所対話構造の認識結果を示す。“Recall”、“Precision”及び“f-measure”は以下の式で求められる。

$$\text{Recall} = \frac{\text{正しく認識した局所対話構造数}}{\text{全局所対話構造数}}$$

$$\text{Precision} = \frac{\text{正しく認識した局所対話構造数}}{\text{認識した局所対話構造数}}$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

実験の結果、局所対話構造の認識率は高くないが、それでも局所対話構造が認識できた分に限っては高い対話行為の認識率を示している。

対話行為の認識結果は洪 [12] の結果を下回った。原因として洪 [12] とは違うコーパスを用いたこと、洪 [12] は学習の際に要素としてそのコーパスに特化したキューフレーズを用いた事が考えられる。

表 1: 対話行為の認識結果 (Recall)

対話名	baseline	決定木	決定木 (直前対話行為利用)	局所対話構造 (Precision)	複合型
A04	42.70%	49.44%	60.67%	46.07% (63.08%)	60.67%
A06	38.10%	51.58%	62.96%	55.56% (88.24%)	66.14%
A21	34.52%	37.06%	44.67%	51.27% (59.06%)	57.87%
average	38.44%	46.03%	56.10%	70.13% (50.97%)	61.56%

表 2: 局所対話構造の認識

対話名	Recall	Precision	F-measure
A04	15.15%	18.52%	0.1667
A06	27.78%	30.30%	0.2899
A21	27.03%	30.30%	0.2857
average	23.58%	26.88%	0.2512

## 4 おわりに

本論文では対話の構造に注目して、発話の対話行為と対話構造を認識する方法を示した。実験により我々の提案方法が有効であることも示した。ただし、局所対話構造の認識率に対しては、あまり良い結果ではなかった。これは使用したコーパス量が少ないということが考えられる。今後、今回のコーパスとは別の領域のコーパスでどの程度有効であるかも確かめる必要がある。

## 謝辞

本研究は科学研究費補助金による。

## 参考文献

- [1] James F. Allen. Recognizing intentions from natural language utterances. *Computational Models of Discourse*, M. Brady & R. C. Berwick (Eds.), Computational models of discourse, pp.107-166, MIT Press, 1983.
- [2] Elizabeth A. Hinkelman, James F. Allen. Two Constraints on Speech Act Ambiguity. *ACL PROCEEDINGS, 27th ANNUAL MEETING*, pp.212-219, 1989.
- [3] 熊本忠彦, 伊藤昭, 海老名毅. 支援対話におけるユーザ発話意図の認識—ユーザ発話文の解析に基づく統計的アプローチ—. *電子情報通信学会論文誌*, Vol.J77-D-2, No.6, pp.1114-1123, 1994
- [4] Masaaki Nagata, Tsuyoshi Morimoto. First steps towards statistical modeling of dialogue to predict the speech act type of the next utterance. *Speech Communication* 15, pp.193-203, 1994.
- [5] Diane J. Litman, James F. Allen. *Discourse Processing and Commonsense Plans*. Phillip R. Cohen, Jerry Morgan, Martha E. Pollack, editors, *Intentions in Communication*, pp.365-388, MIT Press, Cambridge, MA, 1990.
- [6] Arne Jonsson. A model for habitable and efficient dialogue management for natural language interaction. *Natural Language Engineering* 3(2/3), pp.103-122, Cambridge University Press, 1997.
- [7] 高野健治, 島津明. 交通経路案内対話の分析—局所的構造に注目して—. *言語処理学会第10回年次大会*, pp.181-184, 2004.
- [8] Sinclair J. McH., Coulthard R. M. *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press, 1975.
- [9] 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*, Vol.37, No.12, pp.2190-2200, 1996.
- [10] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人. 発話単位タグ標準化案の作成. *人工知能学会誌*, Vol.14, No.2, pp.251-260, 1999.
- [11] Traum, D. R. *A Computational Theory of Grounding in Natural Language Conversation*. Unpublished doctoral dissertation, University of Rochester, 1994.
- [12] 洪陽杓, 白井清昭. 対話行為タグ付きコーパスの作成支援. *言語処理学会第11回年次大会*, pp.815-818, 2005.