CIAIR 車内音声対話コーパスを用いた対話フロー解析

加藤 真吾† 入江 友紀♮ 山口 由紀子‡ 松原 茂樹‡§ 河口 信夫‡§
†名古屋大学工学部電気電子情報工学科 ♮名古屋大学大学院情報科学研究科
‡名古屋大学情報連携基盤センター §名古屋大学統合音響情報研究拠点
gotyan@cogma.org

1 はじめに

実用化された音声対話システム（例えば，カーナビ）の多くは，ユーザの発話に制限が課せられているのが現状である．システムのユーザビリティを高めるために，ユーザの自由な発話に対してシステムが適切に応答できるような方法論が求められる．

本稿では，柔軟性を備えた音声対話システムの対話制御手法の構築を目的として，大量の実データを用いた対話フローの解析について述べる．解析では，CIAIR で作成した意図タグつき対話コーパス [2] を用いる．38 種類存在する意図タグを６つに分類し， 対話フローをオートマトンで表現する．また，対話の構造分析におけるオートマトンの利用可能性について述べる．

2 CIAIR 車内音声対話コーパス

名古屋大学統合音響情報研究拠点（CIAIR) では，実環境における頑健な音声認識の実現や音声対話の高度化を目的として，実走行車内における音声や対話を 1999 年度より収録してきた [1]．このデータベースは，音声，映像，車両操作情報，車両位置といったマルチモーダルな情報が 800 名を超える被験者に対して収録されている．このデータベースに収められているレストラン検索対話は，1999 年度 (対人間)，2000 年度 (対人間，対 WOZ，対システム)，2001 年度 (対人間，対 WOZ，対システム) の計７セッションで構成されている．本稿で解析対象としたのは，2000 年度収録の対人間のセッションであり，その詳細を表 1 に示す．

また，これらの書き起こしデータには，事例を用いた発話意図推定を行うために設計された，発話の意図を階層化して表した意図タグが付与されている [2]．本研究では，この意図タグの第一から第三レイヤまでを利用している．第三レイヤまでの意図タグの種類は 41 種類存在するが，2000 年度収録の対人間のセッションで実際に現れたのは 38 種類であった．表 2 に意図タグつき対話の例を示す．“ D ”はドライバ発話を，“ O ”はオペレータ発話を表す．

3 オートマトンによる対話解析

対話制御を考える上で，まずは実際の対話の構造を定性的に把握する事が望まれる．そこで本研究では，実対話の構造を表現する方法として，解析結果に基づくオートマトンモデルを提案する．

表 1: 解析対象の対話コーパス

被験者数	291
対話数	789
最短ターン数	2
最長ターン数	51
平均ターン数	11.61
意図タグの種類	38

表 2: 意図タグとその発話例

発話例	話者	意図タグ		
		第一（談話行為）	第二 (動作)	第三 (対象)
辛目のマーボー豆腐が食べたいんだけど	D	依頼	検索	店
大黒天横浜楼ピカイチがございます	O	陳述	提示	検索結果
大黒天は駐車場ある	D	依頼	提示	店情報
はいございます	O	陳述	提示	店情報
値段はいくらかな	D	依頼	提示	店情報
五百円からです	O	陳述	提示	店情報
じゃあそこにします	D	陳述	選択	店
大黒天に案内します	O	表明	案内	店
はい	D	陳述	提示	意思内容

3.1 意図タグの分類

対象としたコーパスにおいて意図タグは 38 種類存在しており，そのまま用いたのでは表記が複雑になる．そこで，オートマトン表記の準備段階として意図タグを分類することが望ましい．コーパスのうち約 50 タスクを詳細に観察したところ，ドライバ発話とオペレータ発話が対を構成していることを確認した．また，タスク内にほぼ確実に表れる意図 (検索依頼，情報提示，選択，案内) が存在していることがわかった．本稿では，これらを主発話と呼ぶことにする．また，これらのうち，検索依頼と情報提示，選択と案内がそれぞれ対になっている．そこでまずこれらを記号 (順に I,T,S,H) で表現し，これらとほぼ同等の意味をもつ意図タグをまとめた．次に，対話の調子を整えるあいづちのような役割を果たす意図タグを A，また，I,T,S,H,A のどれにも属さない意図タ

Dialogue Flow Analysis using CIAIR In-Car Spoken Dialogue Corpus : Shingo KATO, Yukiko YAMAGUCHI, Shigeki MATSUBARA, Nobuo KAWAGUCHI (Nagoya University)

表 3: 意図タグの分類

記号	意図	分類された意図タグ	発話例
I(15%)	検索依頼	依頼+検索+店 (駐車場), 依頼+再検索+店 (駐車場), 示唆+検索+店 (駐車場)	中華料理を食べたいな
T(23%)	情報提示	陳述+提示+検索結果, 陳述+提示+駐車場情報, 陳述+提示+店情報	中華飯店が近くにあります
S(15%)	選択	陳述+選択+店 (駐車場), 依頼+案内+店 (駐車場)	じゃあその店をお願い
H(9%)	案内	表明+案内+店 (駐車場)	ではご案内します
A(15%)	あいづち	陳述+提示+意思内容	はい
N(23%)	その他	上記以外のタグ	どういった店がよいですか

グを N とし，最終的に 6 種類に分類した．表 3 にその分類と 789 対話中の出現割合，及び，発話例を示す．

3.2 対話のオートマトンモデル

3.1 節で記号化した 6 種類のタグを用いて，50 タスクの対話データをもとに意図の遷移を表現するオートマトンを人手で構成した．そのオートマトンを図 1 に示す．状態 d1 は開始状態，状態 d2 は入力 T の発話が行われた後の状態であり，それぞれ次発話にドライバの発話を必要とする．状態 o1 は入力 I の発話が行われた後の状態，状態 o2 は入力 S の発話が行われた後の状態であり，それぞれ次発話にオペレータの発話を必要とする．状態 fi は入力 H の発話が行われた後の状態であり，この状態に達すると対話は終了する．状態 su は入力 N の発話が行われたときに遷移する状態であり，タスク中に確実に現れるわけではなく，また，遷移全てを記述すると表記が複雑になるため一つの状態として表している．

コーパスの全 789 対話を用いて，このオートマトンの表現能力を検証したところ，最終状態に達したのは 573 対話[1]であった．その遷移の様子を図 1 に併せて表記する．この遷移の様子から次のことがわかる．

- 789 対話のうち，受理しなかった 205 対話においては，状態 d1 や d2 では次発話にドライバ発話が必要であるにも関わらずオペレータ発話が行われたり，状態 o1 や o2 でドライバ発話が行われたりするなど，同一話者が連続して発話を行うため，一発話で話者が交代しない場合がある．これは，オペレータが検索結果を提示し，続いてその提示した店に関する情報をドライバに伝えるといったような，複数の発話で一連の意図を伝えるときなどに起こる．

- 状態 o1 や状態 o2 から状態 su への遷移，またその逆の遷移が多い．このことより，入力 I と T の対や入力 S と H の対のように発話単位の連接として発話が行われるのではなく，対の間に他の対が出現する場合もある (例えば，I → N(オペレータ発話) → N(ドライバ発話) → T のような場合).

[1]話者交代が正しく行われず，同一話者が連続して発話を行っている 205 対話及び H が出現しない 11 対話を除く

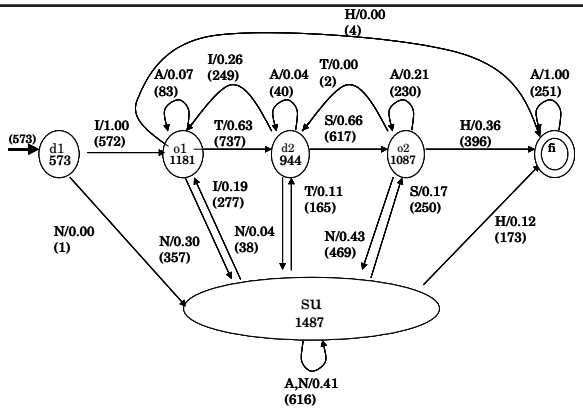

図 1: 対話のオートマトンモデル

- 状態 d1 での入力は I がほとんどであり，状態 o1 での入力は T か N，状態 d2 では I か S，状態 o1 では H か N であるといったように各状態において行われる発話はある程度決まっている．

- 状態 d1 における N の発話数は 1 であり，状態 d1 における総発話の 0.2%，また，状態 d2 での N の発話数は 38 であり，状態 d2 における総発話の 4.0%である．このことよりドライバは主発話が，オペレータは N の発話が多い．

このように，対話フローをオートマトンで表すことにより，対話の開始や終了にはどのような意図の発話が行われるのか，どの状態のときに対話は複雑さを増すのか，ある時点での発話がどの状態から発したものなのかなど，ドメインの特徴や構造を観察することができる．また，対話の流れも直感的に理解しやすく，得られた特徴や構造を用いて対話制御の難しさを軽減できる可能性がある．

4 おわりに

本稿では，CIAIR 車内音声対話コーパスを用いて，対話フロー解析を行った．また，対話をオートマトンでモデル化するとともにその有用性について述べた．発話は対話の状態に基づいて行われており，対話の構造的な性質が明らかになった．

参考文献

[1] 河口，松原，山口，武田，板倉：CIAIR 実走行車内音声データベース，情処研報，SIG-SLP-49，pp.139-144 (2003).

[2] Irie, Y., Kawaguchi, N., Matsubara, S., Kishida, I., Yamaguch i, Y., Takeda, K., Itakura, F., Inagaki, Y.:“ An Advanced Japanese Speech Corpus For In-car Spoken Dialogue Resarch ”, Proceedings of Oriental COCOSDA 2003, pp.209-216 (2003).