

NTTオープンハウス×未来想論2008
「コミュニケーションを理解する：
分析手法, 技術, 未来へのビジョン」
080530@NTT CS研

多人数 × マルチモーダル
インタラクションの分析へ向けて
—音声言語処理技術の拡張の観点から—

高梨 克也

京都大学学術情報メディアセンター
takanasi@ar.media.kyoto-u.ac.jp

言語・音声・対話処理技術の拡張

- 言語処理:
 - 形態素→統語→談話
 - 書き言葉→話し言葉
- 音声認識:
 - 単語→大語彙連続認識
 - 読み上げ→自発音声

書き手・話し手
が対象

- 対話システム:
 - 課題遂行対話→自由会話
 - 二者会話→三者以上(多人数)

→多人数×
マルチモーダルへ

→ボトムアップ(内から外へ): 粒度が小さく, well-formed
で明確な対象(音素, 形態素)の処理を確立し, これを
徐々に**拡張**.

⇔外から中へ(坊農発表): 自然で雑多な状況から対象と
なる現象に**焦点化**(センサリングなど)

多人数 × マルチモーダルインタラクション

- 多人数・時間同期型コミュニケーションはマルチモーダルになるのが一般的.

※複数人での音声電話？

- マルチモーダルな二者会話は普通.

マルチモーダル



多人数

- 分析上の課題:

① 多人数: 受け手・次話者の決定, 参与役割の割り当て

② マルチモーダル: 振る舞い間の連鎖性の明確さ

分析上の課題①: 多人数

受け手・次話者の決定, 参与役割の割り当て
「誰が・誰に」

- 多人数(3人以上)の会話＝聞き手が2人以上
→ 2人(以上)の聞き手の間の区別:
 - 受け手・次話者
 - 傍参与者・非次話者
 - 受け手/次話者選択のための方法:
 - 非言語: 視線, 頭部, 身体方向など(大塚発表)
 - 言語: 呼びかけ(明示的), 共有知識や会話文脈(非明示的)
- 言語的 & 非明示的な選択の分析は不十分

三者テレビ会話（視線は合わない）の例

■ Ex1

[CがAとBに「ゼミはもう決まっているのか」と質問]

1 B あ、まだですね. 入ってからだよね.

丁寧体

非丁寧体

初対面Cへの回答

友人Aへの確認

■ Ex2

1 Aカタクどうなった? ⇒ 「カタク」にまつわるエピソードを共有しているBへの質問

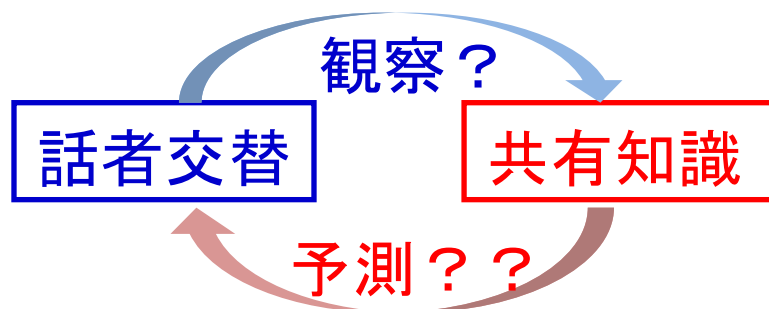
2 B あれはな一部 ⇒ Aへの応答

3 A ああ

4 B キャンプでさああだ名が決定してん

⇒このエピソードを知らないCへの説明

5 C ああ



分析上の課題②: マルチモーダル 振る舞い間の連鎖性の明確さ 「何を・どのように」

- 言語的発話の間関係の例: 隣接ペア

第一部分

第二部分

「質問」

→

「応答」

「依頼」

→

「受諾/拒否」

→ 条件的関連性: 明確な連鎖性

– 第二部分が規則的に生起

– 第二部分の不在も参与者に気付かれる

- 非言語行動の場合: 緩やかな連鎖性

– ジェスチャーや表情が他の参与者の振る舞いに影響する仕方は言語的発話の場合と同じか?

発話・視線・指差し・うなずきの「連鎖」???

■ Ex3

A: (F_えっと)|まず

(N_河原)研究室では

音声言語処理を|用いた

デジタルアーカイブの|高度化というの|目標に

|研究を|行なっています

[Video](#)

→非言語行動の「不在」は顕在化するか？

技術的応用の方向性の比較

- 対話に参加する:
 - 対話システム(人間-コンピュータ間インタラクション)
 - ユーザ発話を正しく理解し, 適切な応答を生成することが必要.
 - ユーザがシステムの能力に適用してくれる面も.
- 人間同士のコミュニケーションを支援する:
 - 要約, 議事録作成, CSCW, 情報可視化
 - 会話能力(発話理解・生成)は完全でなくてもよいかも.
 - 人間同士の会話についていく能力が必要.
(システムの能力への「手加減」がなくなる)

※「正解率」という評価方法の問題

- 評価対象となる単位の正解が明確.
 - 当該技術のモジュールとしての応用価値が自明.
- 応用の文脈での評価: 顧客満足度, リピーター, 売上?