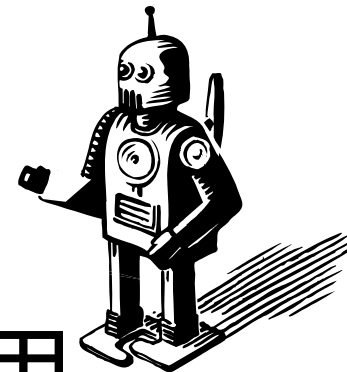
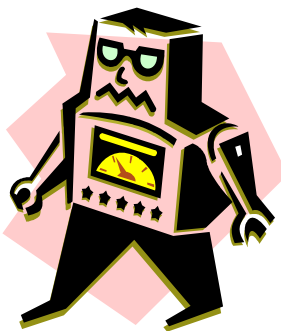


【2010年3月11日 言語処理学会第16回年次大会】



NTT



対話システムのための
「なりきり質問応答」を用いた
質問応答ペアの収集とその応用

東中竜一郎* 堂坂浩二** 磯崎秀樹**

*NTTサイバースペース研究所

**NTTコミュニケーション科学基礎研究所

目的

- 雑談システム実現のため、対話システムのための発話ペアを大量かつ効率的に収集
- 発話ペア:



- 個性の実現やタスク外入力への対応のためシステム毎に発話ペアの作成は必須
- 発話ペアが十分網羅的であれば、システムの応答能力が向上し、システムの満足度が向上

A.L.I.C.E. (レブナー賞受賞)の発話ペア

○ HAVE YOU SEEN BLADE RUNNER

○ Sure I have seen it.

But Replicants aren't Robots.

○ YOU SOUND LIKE HAL

○ To me that's a great compliment.

A.L.I.C.E.は約47000のパターンを使用

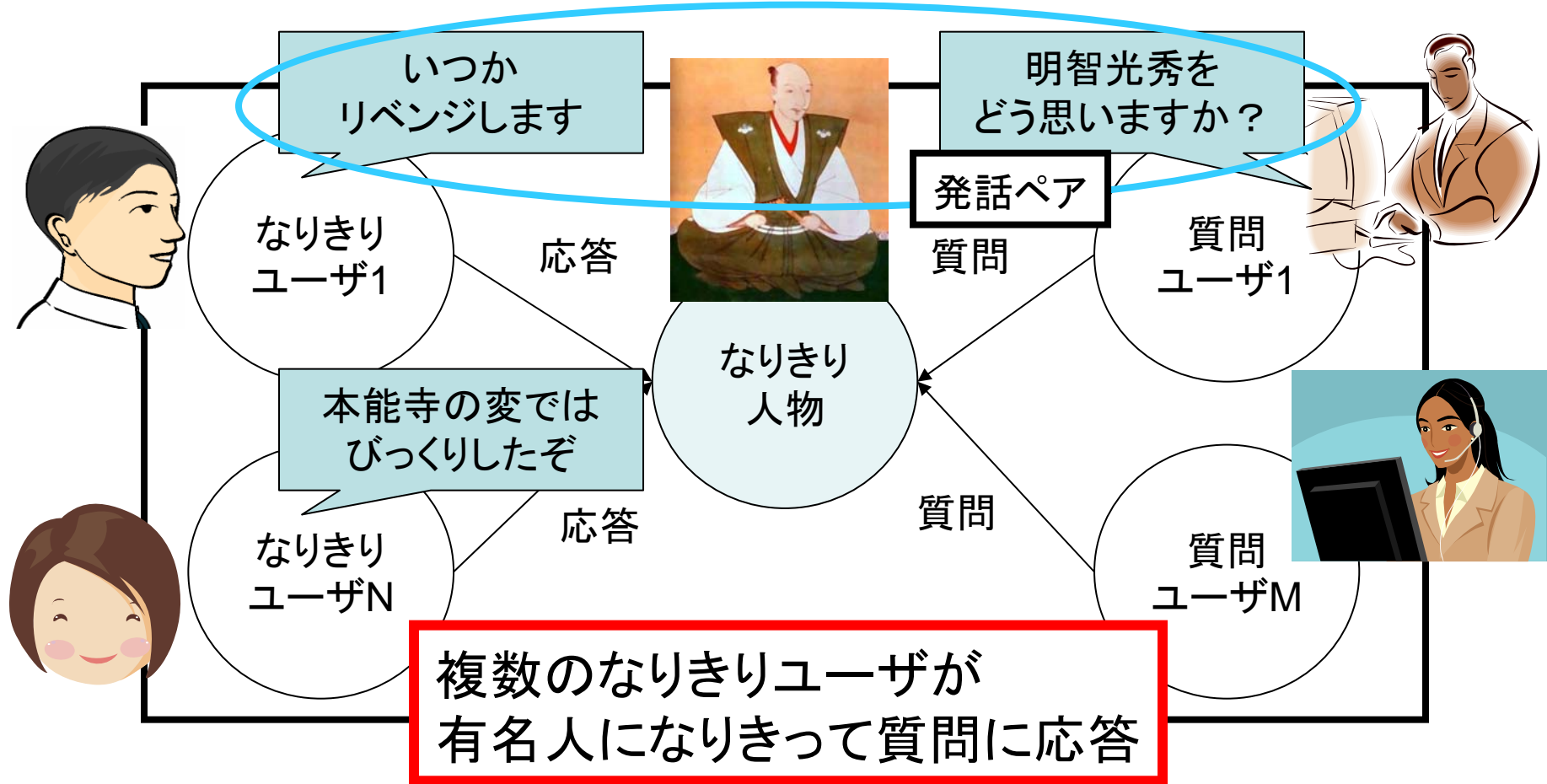
その他, 国内の音声対話システム「たけまるくん」は
6000の入力に対し300の応答を利用(竹内ら, 2007)

問題

- 手作業による発話ペアの作成
 - 大量のペアの準備が必要で高コスト
 - 基本的にシステムごとに一から作成する必要
- 掲示板からの発話ペアの自動獲得
 - ノイズが多い
 - 不特定多数のやり取りが対象
 - ⇒ 発話ペアが矛盾を含む・個性が一貫しない
- 個人のやり取りを大量に収集
 - プライバシーの問題
 - 絶対量がどうしても少ない・個人に高負担

アプローチ:「なりきり質問応答」

- 有名人になりきって質問応答を行うことを趣旨としたサイトで発話ペアを効果的に収集



なりきり質問応答の利点

- ユーザの能動的参加の促進
 - なりきる楽しさ→答えたくなる
 - 自分の好きな人物に質問可能→質問したくなる

<楽しさはデータ収集では重要> (von Ahnら)
- 複数人で1人物の応答を肩代わり
 - 個人の応答負担を軽減
 - 迅速な応答
 - 効率的な質問応答サイクルの実現
- プライバシーの問題が比較的少ない

データ収集実験

- なりきり質問応答サイトを構築
 - 初期状態として4人物・12質問・36応答を登録
- 実験参加者
 - 50名(男女同数)・自宅から参加
 - 謝礼あり(一定額を支給)
- 教示
 - 期間内に100の投稿(質問または応答)を行う
 - 質問・応答はなるべく同数にする
 - 自演応答はカウントに含まない
- 期間:3週間

なりきり質問応答サイト トップ画面



user8さん、こんにちは。

[トップページ](#)

ようこそ

ユーザ／サイト全体の投稿数

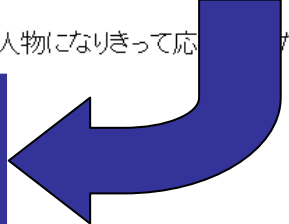
なりきり質問応答サイトへようこそ。みんなが知っている人物に質問したり、質問に対してその人物になりきって応答してください。

あなたの現在までのデータ

あなたの質問回数	あなたのなりきり応答回数	あなたの書き込みの総数
21	99	120

サイト全体のデータ

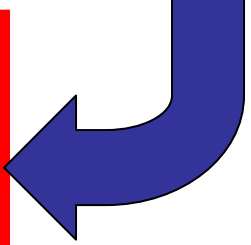
質問総数	なりきり応答総数	書き込み総数
2497	3834	6331



特定の人物で会話しましょう

ユーザが登録した人物の一覧

- 会話したい人物をクリックして、質問応答を始めてください。
- 中川翔子(しょこたん)(13) / 紀貫之(2) / ぼのぼの(1) / デビルマン(3) / モーツァルト(12) / 司馬遼太郎(3) / ベートーヴェン(6) / 上杉謙信(3) / トム・クルーズ(5) / たもり(5) / マイルス・デイビス(2) / 吉田拓郎(2) / 篤姫(3) / 美空ひばり(6) / 木村拓也(4) / チンギスハーン(4) / シーザー(9) / 森繁久彌(3) / モロボン・ダン(4) / 松坂大輔(16) / 金大中(3) / 清少納言(8) / サンタクローズ(20) / レオナルド・ダ・ヴィンチ(10) / かぐや姫(16) / ジジ(魔女の宅急便)(7) / 伊集院光(5) / スザンヌ(8) / キティー(16) / シェア・アズナブル(9) / ガチャピン(14) / 安室奈美恵(6) / アンパンマン(24) / ミッキー・マウス(20) / ネス(MOTHER2)(2) / 黒木瞳(10) / 夏目漱石(16) / 青木さやか(4) / 福田康夫(14) / どーもくん(3) / 藤原紀香(29) / 綿矢りさ(2) / 東国原英夫(13) / たねぼんだ(10) / 西村ひろゆき(3) / 林家パー子(6) / 細木数子(7) / ちびまる子(11) / うーたん(NHK教育「いないいないばあっ!」)(4) / マリーアントワネット(8) / 蒼井優(7) / ルイージ(7) / IKKO(7) / チュートリアル(5) / キムジョンイル(9) / 宮崎あおい(7) / 泉こなた(らき☆すた)(5) / アムロ・レイ(8) / タケカワ ユキヒデ(4) / 寅さん(13) / 坂本金八(6) / 山田太郎(ドカベン)(7) / 中居正広(16) / 萩尾望都(4) / 手塚治(9) / 野村克也(15) / 斎藤隆(2) / 大沢啓二(4) / 唐川侑己(2) / 赤川次郎(3) / ゲゲゲの鬼太郎(9) / つげ義春(4) / 卑弥呼(8) / チョッパー(ワンピース)(3) / ルフィ(ワンピース)(4) / ジョン・ウェイン(2) / 古今亭志ん朝(5) / 紫式部(9) / 徒然亭若狭(5) / ガンジー(6) / チャーチル(7) / ダルビッシュ有(7) / 島田伸介(8) / ヘレン・ケラー(1) / 石川啄木(7) / 本居宣長(3) / カブ(3) / 夜神月(7) / 面堂終太郎(2) / 明智小五郎(5) / ジャイアント馬場(10) / 安倍晋三(12) / シェーン(4) / スターリン(3) / スティーブ・マッキーン(3) / オノ・ヨーコ(6) / 杉田かおる(4) / 安藤美姫(17) / 松井秀喜(20) / ロナウジーニョ(8) / イチロー(21) / 西行(6) / 菅原文太(3) / 上村愛子(5) / 圭一(1) / ハッハ(10) / 小島よしお(22) / アントニオ猪木(3) / 天沢聖司(百をすませば)(3) / ラムちゃん(うる星やつら)(7) / 沢尻エリカ(12) / おじゃる丸(6) / 平井堅(10) / 涼宮ハルヒ(7) / 室伏広治(8) / アグネス・チャン



さあ、織田信長になりきって応答しましょう

織田信長への質問(新着順) > [すべてを見る\(6\)](#)

織田信長への質問

質問No:2000 [・貴方が登場する小説・漫画・ゲームが色々ありますが、一番のお気に入りは何ですか？](#)

応答・パチンコだね。

なりきり応答入力 なりきり応答をしてください。文章1行(100文字まで)をお願いします。

質問

織田信長への質問

質問No:773 [・明智光秀に対して一言ください。](#)

応答・よくやってくれたよ。もう日本統一ば面倒くさくなってたからネ。

応答・近いうちにリベンジに行くぞ。

なりきり応答入力 なりきり応答をしてください。文章1行(100文字まで)をお願いします。

**ユーザの
なりきり応答**

織田信長への質問

質問No:715 [・本能寺の変がなかったら、どんなふう to 日本を変えていますか](#)

応答・王国にして、ルイ14世並みに繁栄させたな。

なりきり応答入力 なりきり応答をしてください。文章1行(100文字まで)をお願いします。

織田信長への質問

質問No:683 [・ライバルは誰ですか？](#)

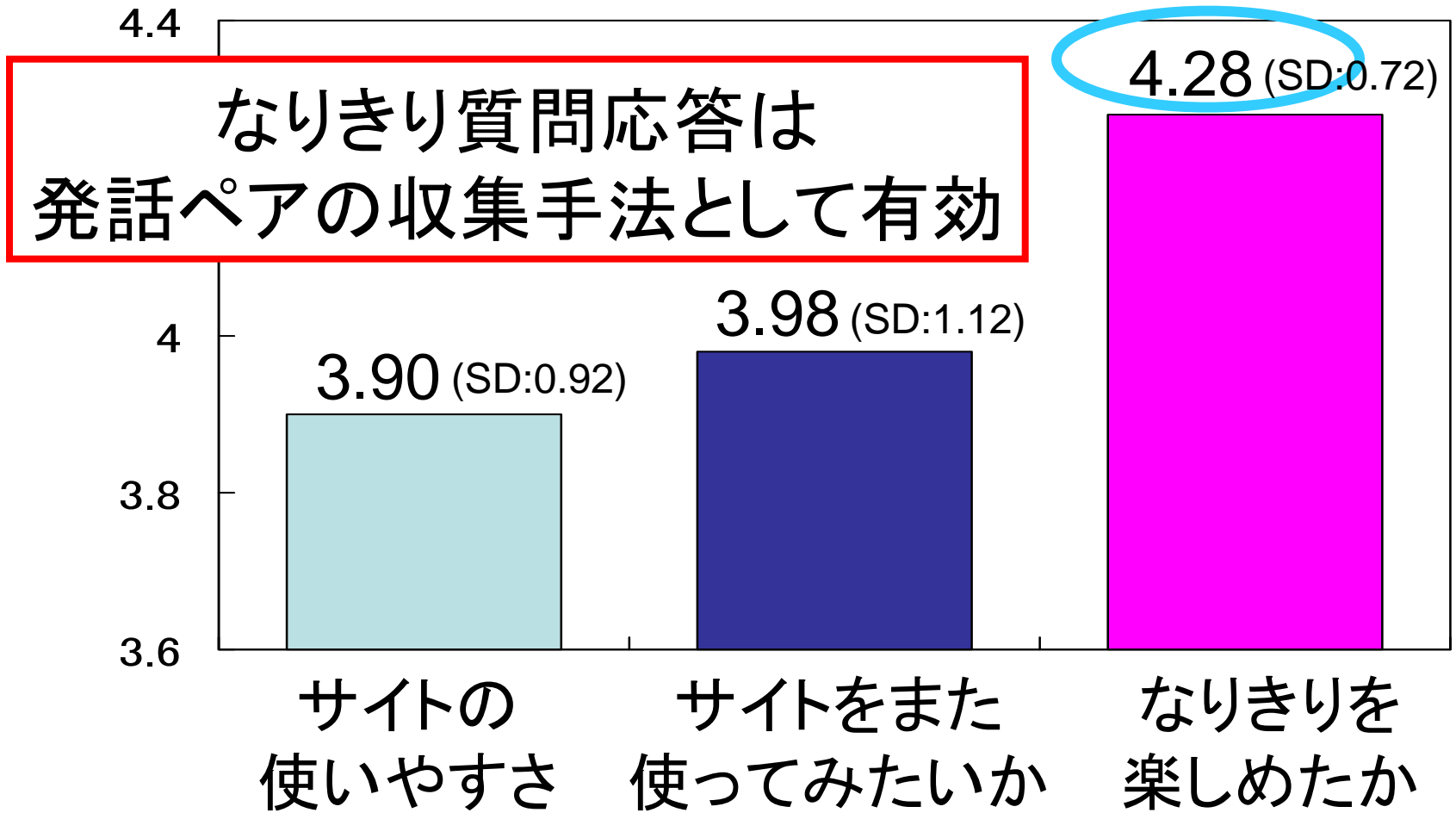
応答・妻夫木君に決まってるジャン。

結果：収集されたデータの概要

登録された人物数	397
質問数	2502
回答数	3838
投稿の総数	6340
自演応答数	213
150以上投稿したユーザ数	9
200以上投稿したユーザ数	3
1質問に回答したユーザ数	1.62
1人物あたりの回答ユーザ数	5.21

- 規定数を超える投稿数
- 1人物のやり取りを複数のユーザ(低負担)で実現

評価：50ユーザの主観評価（5段階）



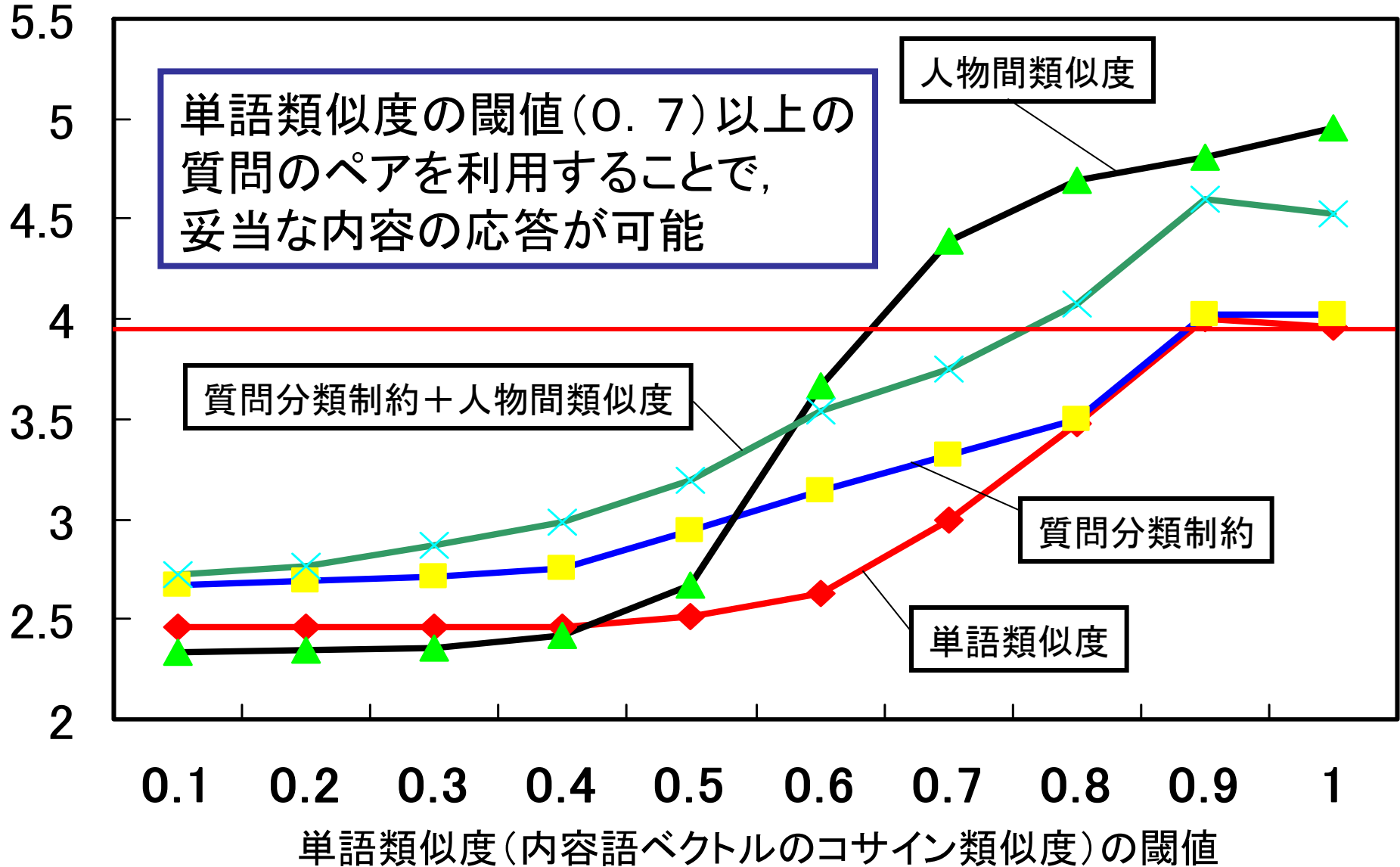
なりきる楽しさがデータ収集に有効に機能

おまけ：収集した発話ペアの応用

- 有名人対話システムの作成
 - 各有名人に対して得られた発話ペアを用いて、その人物らしく応答する対話システムが実現可能
- 問題
 - 人物によっては発話ペアが少なく応答に限界
- 解決策
 - 他の人物の発話ペアを利用して応答
 - 4手法(単語類似度, 質問分類制約, 人物間距離, 質問分類制約+人物間類似度)を提案
 - 入力質問と類似する質問を持つ発話ペアを検索
- 他人物の発話ペアを用いた質問応答実験を実施

評価結果：応答内容の適切さ

(7段階評価, 967質問について評価)



4つの相互利用手法の出力例



質問対象:ビル・クリントン
質問:「影武者はいますか？」

それは答えられないな。(オリジナル回答)

単語類似度



○危なくなったときに、かくまってくれそうな人なら何人か。

質問分類制約



○えー作るヒマないですよおー

人物間類似度



○決して油断できぬヤツだな

質問分類制約
+ 人物間類似度



○敵が多すぎて特定は出来ない。
尊敬するリンカーンの末席に並べたのが幸せ。

4つの相互利用手法の出力例



質問対象:ビル・クリントン
質問:「影武者はいますか？」

それは答えられないな。(オリジナル回答)

単語類似度

杉田かおる

- 仲のいい友達はいますか？
- 危なくなったときに、かくまってくれそうな人なら何人か。

質問分類制約

安藤美姫

- 彼氏はいますか？
- えー作るヒマないですよおー

人物間類似度

チャーチル

- ルーズベルトをどう思いますか？
- 決して油断できぬヤツだな

質問分類制約
+ 人物間類似度

J. F. ケネディ

- あなたを撃った人間に心当たりはありますか・・・？
- 敵が多すぎて特定は出来ない。
尊敬するリンカーンの末席に並べたのが幸せ。

まとめと今後の課題

- 対話システムの発話ペアの収集法として「なりきり質問応答」を提案
 - ⇒ データ収集実験によって有効性を検証
- 発話ペアの再利用手法を提案
 - ⇒ 評価実験により一定の有効性を検証
- 今後の課題
 - 収集された発話ペアの質の確保
 - 発話ペアの実際の対話システムへの応用
 - 大規模データ収集実験の実施