

ボット名：**MarikoZatsudanBot**

カスタマイズ可能な オープンドメイン雑談対話エンジンの開発

角森唯子 大西可奈子 藤本拓
角野公亮 吉村健 磯田佳徳
(NTTドコモ)

1. 背景
2. 自然対話プラットフォーム
 - シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン
3. 雑談対話エンジン
 - 概要
 - 大規模データベース
 - カスタマイズ機能
4. ライブコンペ
 - チャットボット構築
 - ユーザ評価
5. まとめと今後の課題

1. 背景

2. 自然対話プラットフォーム

- シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン

3. 雑談対話エンジン

- 概要
- 大規模データベース
- カスタマイズ機能

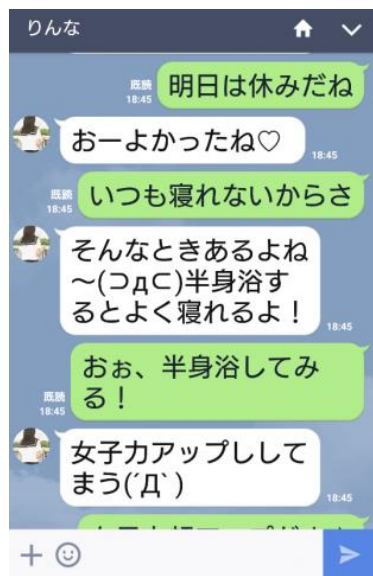
4. ライブコンペ

- チャットボット構築
- ユーザ評価

5. まとめと今後の課題

雑談対話システムの需要の高まり

→ スマホや玩具, ロボットなど, 利用シーンの多様化



<http://atom2020.jp/>



<https://www.takaratomy.co.jp/products/omnibot/ohanas/>

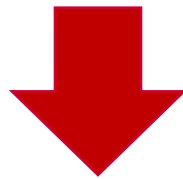
<https://news.microsoft.com/ja-jp/features/big-dreams-behind-xiaoice/>

利用シーンに合わせた対話を行うことで、**ユーザの対話継続欲求**
が高い対話を実現できる可能性が高い

■ 従来の対話システムの課題

- 特定のドメインにしか対応できない
- カスタマイズが困難

→ 利用シーンが限定されてしまう場合が少なくない



カスタマイズ可能なオープンドメイン雑談対話エンジンを開発

- NTTドコモでは「自然対話プラットフォーム」という共通プラットフォームを開発しており、本エンジンはその中で提供
- 自然対話プラットフォームは、シナリオ対話・意図解釈・知識Q&A・雑談対話 の4エンジンで構成

1. 背景

2. 自然対話プラットフォーム

- シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン

3. 雑談対話エンジン

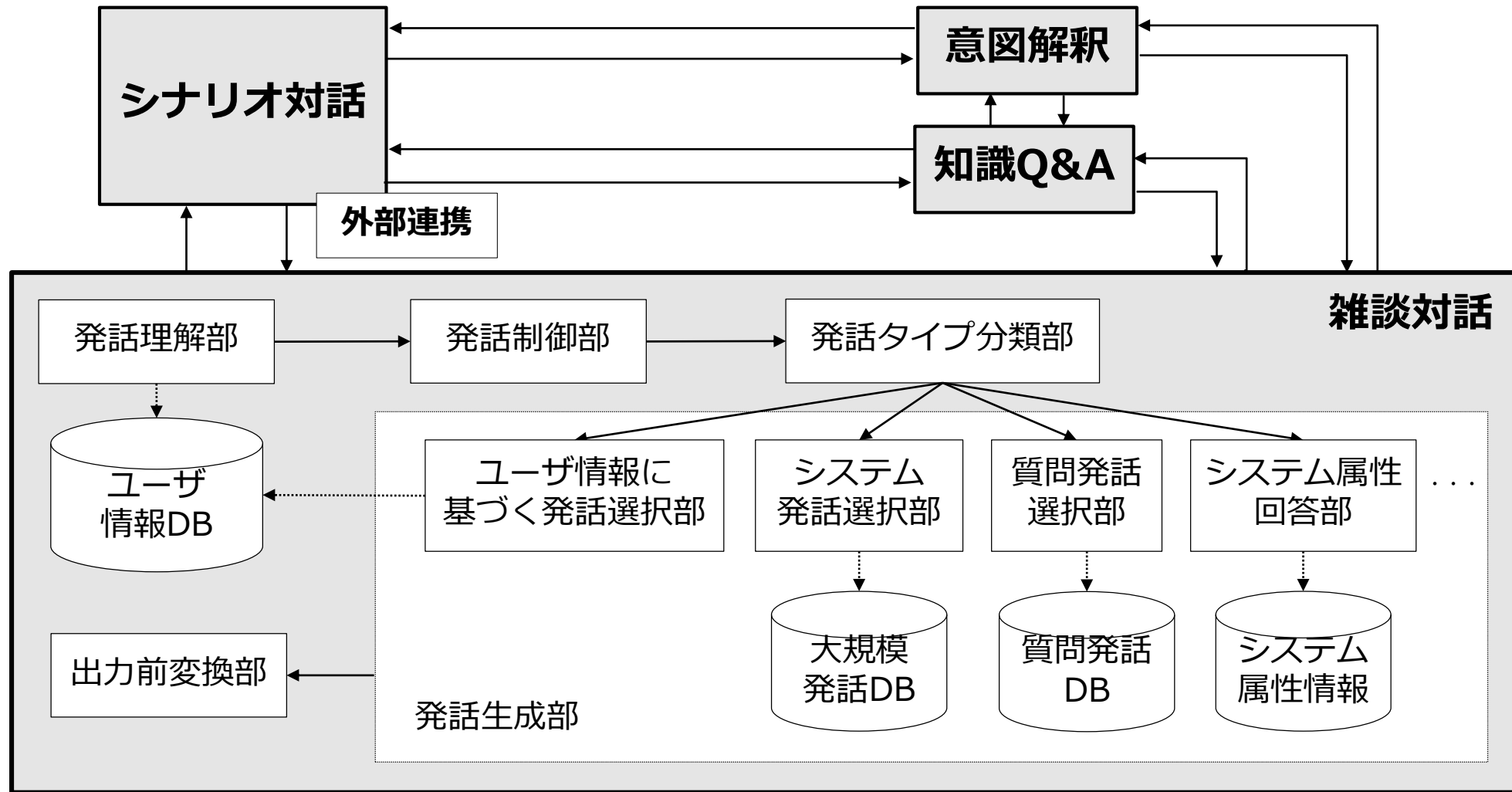
- 概要
- 大規模データベース
- カスタマイズ機能

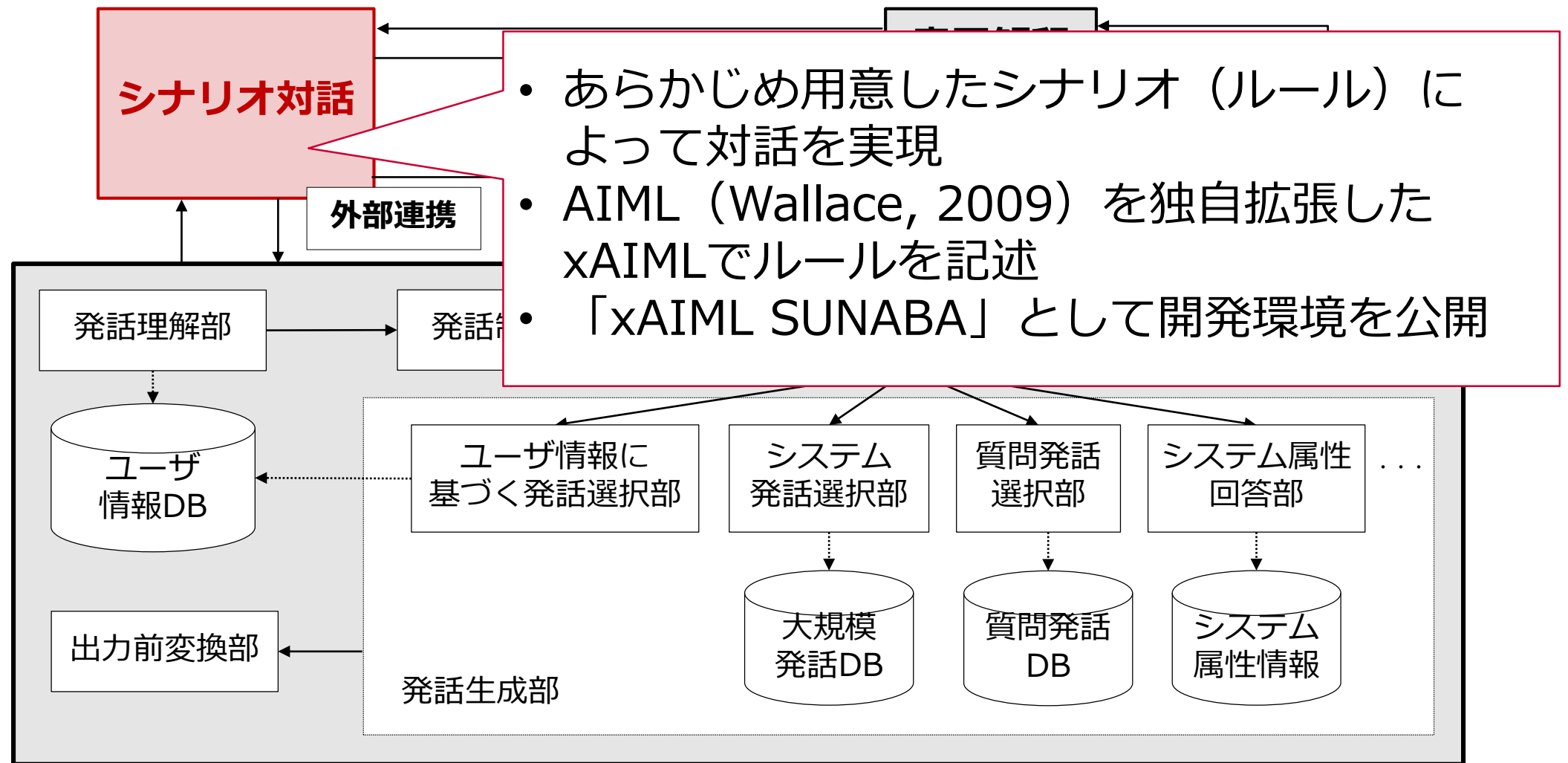
4. ライブコンペ

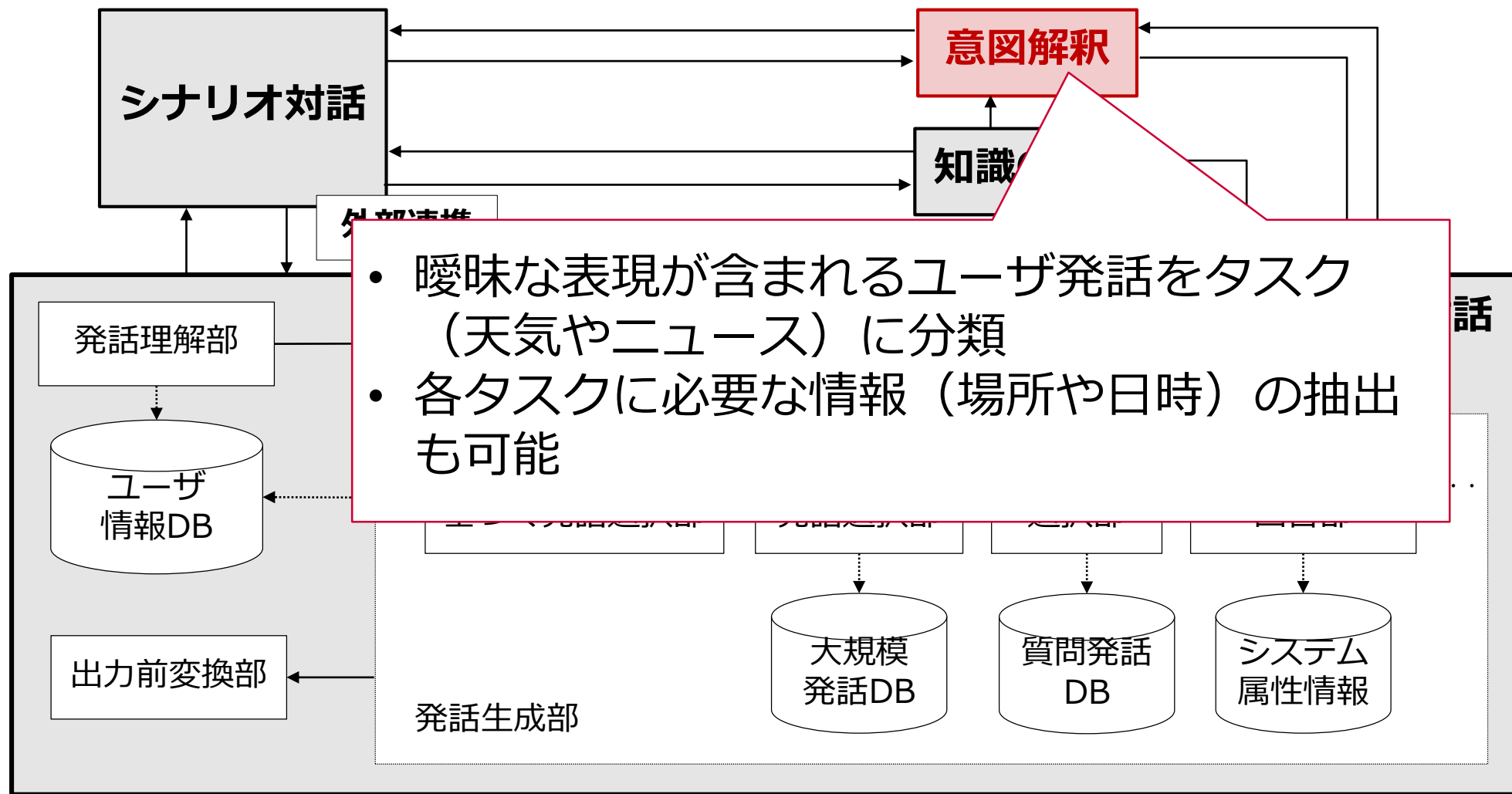
- チャットボット構築
- ユーザ評価

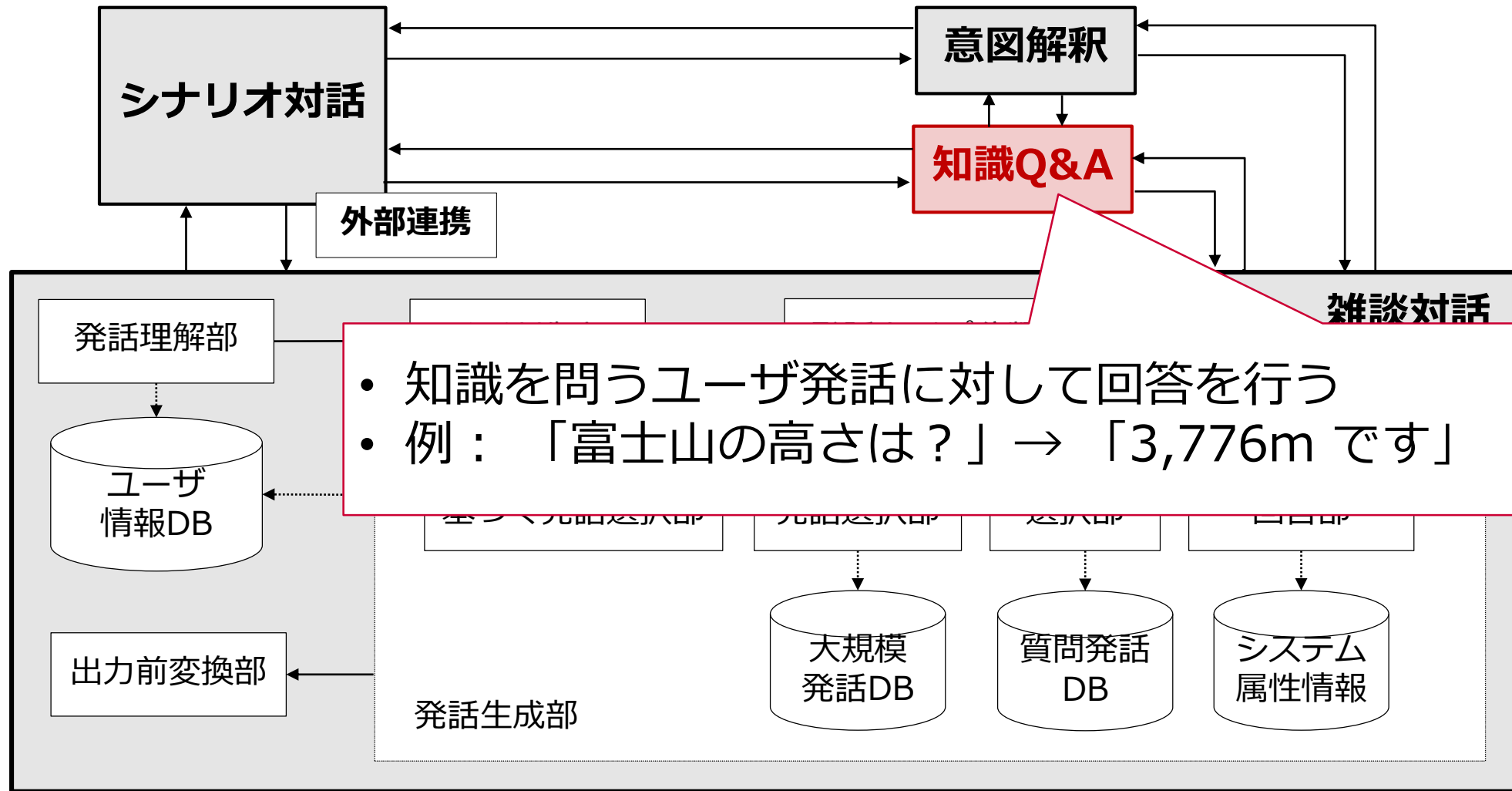
5. まとめと今後の課題

自然対話プラットフォーム

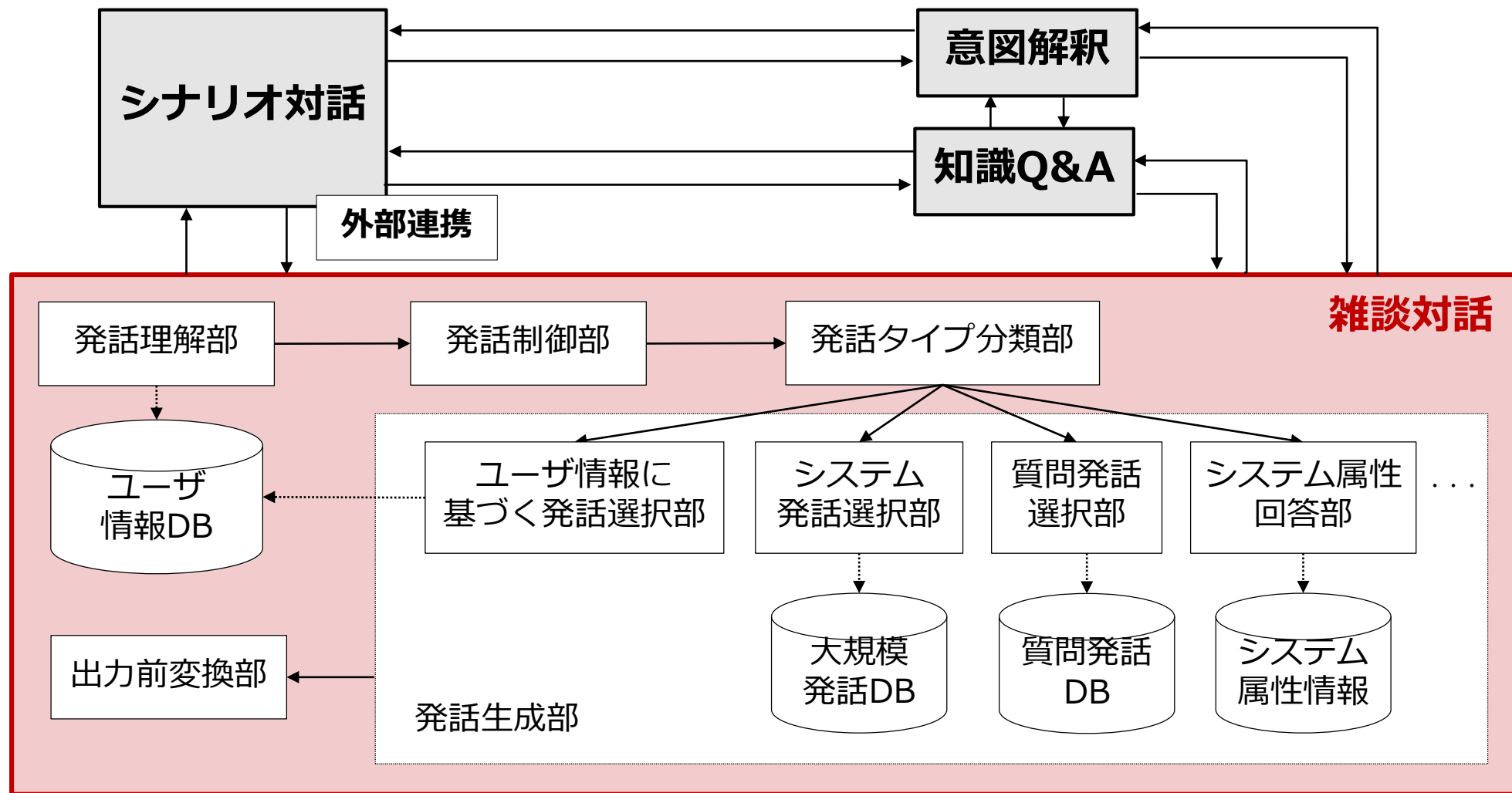








1. 背景
2. 自然対話プラットフォーム
 - シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン
- 3. 雑談対話エンジン**
 - **概要**
 - **大規模データベース**
 - **カスタマイズ機能**
4. ライブコンペ
 - チャットボット構築
 - ユーザ評価
5. まとめと今後の課題

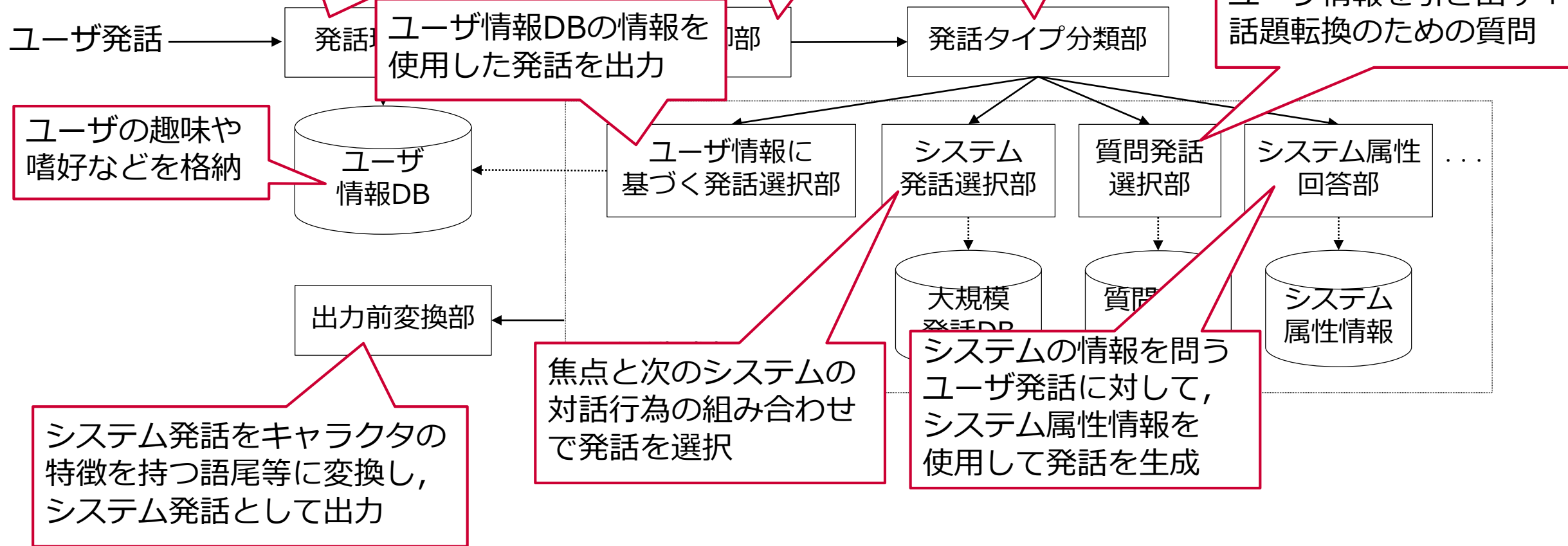


雑談対話エンジンの概要

- ユーザ発話から焦点と対話行為を推定
- ユーザ発話にユーザ自身に関する情報が含まれる場合には抽出

これまでの対話の履歴から、次のシステム発話の対話行為を決定

次のシステム発話を出力すべき発話生成部のモジュールを決定



■ 大規模発話データベース

- Web 上の豊富なトピックを利用して発話することで、多様な話題に対応できる**オープンドメイン**な対話を実現

■ カスタマイズ機能

- 下記のカスタマイズにより、あらゆるシーンで利用可能
 - キャラクタ性の付与
 - シーンにそぐわない不適切発話の除去
 - システム発話選択モジュールの優先度の変更

大規模データから述語項構造から発話DBを生成 (Higashinaka, 2014)

(例)

「先週は今治に行ったんですけど、焼鳥が有名らしくて、とても美味しかったです。」

焦点：今治
述語：美味しい
主語：焼鳥

(対話行為：質問_事実)

「焼鳥が美味しいですね？」

大量のWebデータ

述語項構造解析
(Imamura+, 14)

焦点とそれに紐づく述語項構造のペアを抽出

発話文変換

述語項構造ペアを平叙文に変換し、各対話行為に即した発話文になるよう文末を変換

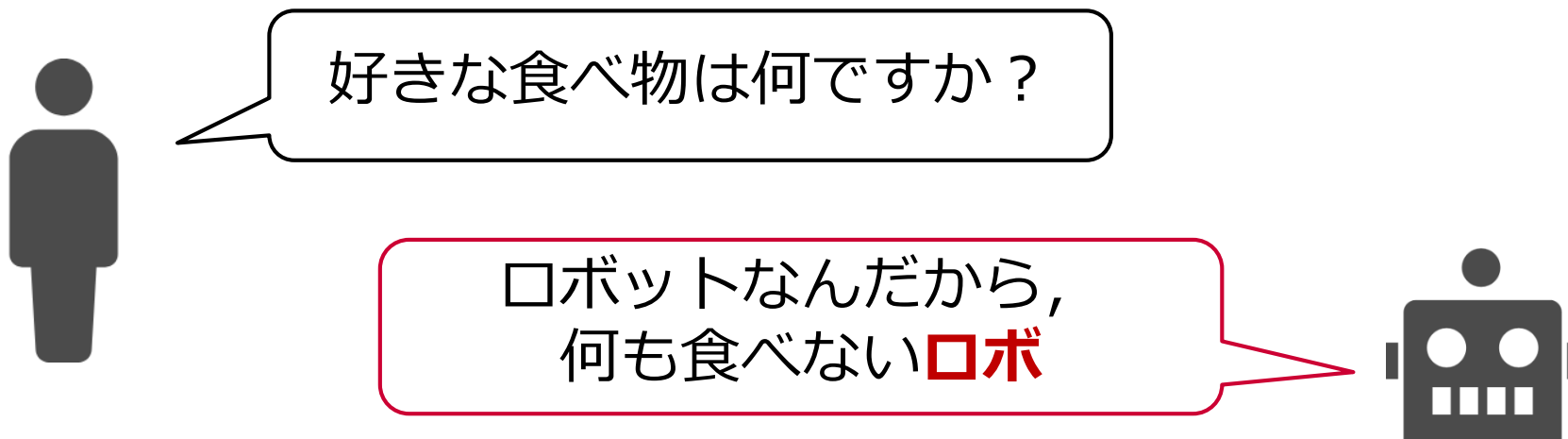
発話文を格納

焦点	発話
今治	焼鳥が美味しいですね？
⋮	⋮

約4,000万発話

■ キャラクタ性の付与

- キャラクタのプロフィール情報に基づく発話が可能
- キャラクタに関する文書から発話文を自動生成することで、よりキャラクタに特化した焦点に応答可能
- 出力前変換部において、キャラクタ風の口調にシステム発話を変換



■ 不適切発話の除去

- 利用シーンによって不適切な発話の定義は異なる
- 焦点とそれに紐づくシステム発話に対して、センシティブ情報（モラル違反/タブー/デリケートなど）のラベルを付与する技術を開発（NTT 協力）
- 利用シーンに応じて削除するラベルを決定し、そのラベルに紐づく発話をデータベースから除去

■ システム発話生成部モジュールの優先度

- 発話タイプ分類部（次のシステム発話を出力すべきモジュールを決定）において、モジュールの優先度を変更することが可能
- 複数のモジュールで選択された発話を連結し、1 システム発話として出力することも可能

1. 背景
2. 自然対話プラットフォーム
 - シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン
3. 雑談対話エンジン
 - 概要
 - 大規模データベース
 - カスタマイズ機能
- 4. ライブコンペ**
 - チャットボット構築**
 - ユーザ評価**
5. まとめと今後の課題

自然対話プラットフォームの技術を用いて、カスタマイズしたチャットボット (**MarikoZatsudanBot**) を構築

- システム主導で対話が展開されるように設計
 - [質問発話選択部]と[ユーザ情報に基づく発話選択部]が選択される優先度を高くした
- 親しみやすいキャラクタの設定
 - 関西弁の女性キャラクタ
- キャラに合わせた不適切発話の除去
 - 政治や宗教, 国家に関する発話は不適切と考え, 発話DBより除去
- シナリオ対話エンジンとQ&A エンジンとも連携している

ライブコンペ予選におけるユーザ評価

約30名のクラウドワーカーによるユーザ評価：「どれくらいまた話したいと思うか」で5段階（1が最良）

■ 評価結果スコア：**1.97**（参加していた11 チーム中で**最も良い**）

■ 対話例

話者	発話
システム	[user name] さん、こんにちは。私の名前はマリコやで。今日はなにしとったん？
ユーザ	今日はイチジクのコンポートを作りましたよ。
システム	赤ワインで作ってるんだね
ユーザ	そうです。庭にたくさんイチジクが実ったんでね。

■ 出現頻度の低い焦点（コンポート）に対しても自然な応答ができており，大規模発話DBが有効に機能

■ 関西弁については，概ね親しみやすいという意見であり，キャラクター性の付与には一定の効果がある

1. 背景
2. 自然対話プラットフォーム
 - シナリオ対話エンジン, 意図解釈エンジン, 知識Q&Aエンジン
3. 雑談対話エンジン
 - 概要
 - 大規模データベース
 - カスタマイズ機能
4. ライブコンペ
 - チャットボット構築
 - ユーザ評価
- 5. まとめと今後の課題**

■ まとめ

- ユーザの対話継続欲求の高い対話の実現を目標とした、利用シーンに合わせてカスタマイズ可能な**オープンドメイン雑談対話エンジン**の開発
- 本エンジンの特徴
 - 大規模応答データベースを保持し、あらゆる入力に対しても応答が可能
 - 利用シーンに合わせたカスタマイズが可能
- ライブコンペ予選のユーザ評価において、ユーザの対話継続欲求が高い対話を実現できていることを確認

■ 今後の課題

- 本エンジンの技術は、商用サービス「かたらい*」で提供されている
- 実サービスで使用される中で課題を抽出していき、より自然な対話を行うエンジンへと改良を続けていく

*かたらい : <https://www.katarai.ai/>



