

# 快適な情報享受を可能とする音声対話システム

高津 弘明<sup>1</sup> 福岡 維新<sup>1</sup> 藤江 真也<sup>1,2</sup> 林 良彦<sup>1</sup> 小林 哲則<sup>1</sup>

早稲田大学<sup>1</sup>

千葉工業大学<sup>2</sup>

{takatsu,fukuoka}@pcl.cs.waseda.ac.jp, shinya.fujie@p.chibakoudai.jp,  
yshk.hayashi@aoni.waseda.jp, koba@waseda.jp

## 1 はじめに

ニュース記事のようなまとまりのある情報を快適に享受する(受け取って自分のものにする、またその行為を楽しむ)ための仕組みを提案し、意図性の異なる多様な情報行動を混在させて情報にアクセスできる会話システムを実現する。

人間の情報に関する行動のうち、情報獲得・享受に関する行動は、意図性の有無によって、意図的な情報探索と意図性のない受動的な情報行動に分類できる [1]。コンピュータサイエンスの領域では、主に前者に焦点が当てられ、情報検索のような分野で活発に研究されてきた。一方、受動的な情報行動は、偶発的な状況によることが多いため、研究要素に乏しいと考えられてきた。しかしながら、我々が日常で行う情報行動は両者の区別が明確ではなく、多様な意図性をもつ行動が複合したものである。

特に、ここで対象とするニュース記事のようなまとまった量の情報を得る場合、このような複合した情報行動の扱いは必須となる。例えば、ニュースをラジオで聴く行為は、受動的な情報行動の典型例である。ラジオの聴取では、リスナーは受け身で聞いていればよいので楽に情報を取得できるという長所がある。しかし、途中で分からない単語が出てきても質問ができない、聞き逃した内容を聞き返せない、興味がない内容であっても最後まで聴き続けなければならないという短所がある。一方、対極にあるのが質問応答型の対話システムである。能動的な情報行動である質問応答は、断片的な情報を得るのには適しているが、まとまった量の情報を得ようとすると、次々と質問を考えながら質問し続けなければならないので現実的でない。つまり、快適な情報享受は、質問応答のような意図性の高い能動的な情報行動だけでも、聞き流しのような意図性のない受動的な情報行動だけでも実現できず、これらの状態を自由に行き来することにより達成されると考えられる。

本研究で提案する音声対話システムでは、利用者は基本的には受け身の姿勢で、要約・口語化された内容を聞くが、理解や興味の状況に応じて、意識的・無意識的な情報要求により、システムから必要十分な情報を得ることができる。ここで、意識的な情報要求とは明示的な質問や感想などの意図性の高い反応を示し、無意識的な情報要求とは「え?」や「へー」などの感嘆表現や表情などの意図性の低い反応を示す。このような無意識に発した感嘆表現や表情が情報要求となる現象は、音声対話特有の現象であり、これに即応性高く適切に応答することが情報行動の活性化と情報享受の質の向上につながることを考える。そこで、要約で省かれた情報や予測される質問に対する回答をあらかじめ発話計画に組み込んでおくことで、このような即時的な情報要求に対処する。

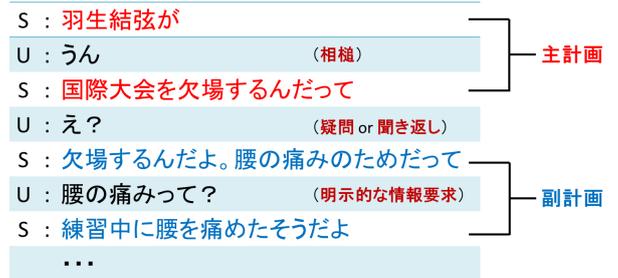


図 1: 想定する会話例

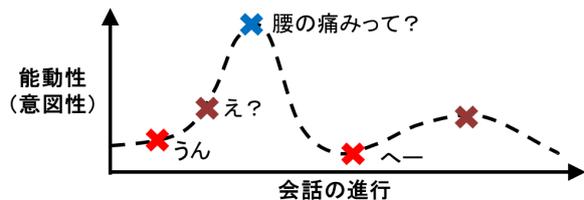


図 2: 会話の進行に伴う利用者の意図性の変化

以下、2章で関連研究について述べ、3章で本研究で想定する会話の例を示す。4章でシステムの要件とアーキテクチャについて解説し、5章と6章で発話計画の生成方法について説明する。最後に、7章で音声対話システム上での実装について述べる。

## 2 関連研究

古くから、ユーザーの発話に応じて情報を提示する質問応答型の対話システムが研究されてきた [4, 5]。これらのシステムでは、ユーザーの明示的な質問に対して断片的な情報を提供するのにとどまっていた。近年では、質問応答機能を持ち、システムが主体的に情報を提供する対話システムの研究が進み [6, 7]、まとまった量の情報提供を行う音声対話システムも提案されている [8, 9]。これらのシステムでは、受動的な情報行動と能動的な情報行動の両方を扱ってはいるものの、システムが情報を提示している間は、ユーザーは黙って聴いていなくてはならない。また、ひと通り内容を伝え終えた後は、一問一答形式で質問を受け付けるモードに移行し、ユーザーからの質問を待機することになるが、それにより会話のリズムが失われてしまう。このように、ユーザーが反応を示せるタイミングは限定的であり、音声対話本来の快適性を十分に活用できていないとも言えない。また、これらのシステムでは、元の情報をそのまま伝えていることが多く、会話インタラクションでの伝達を意識した情報の抽出や発話文の書き換えは行われていない。

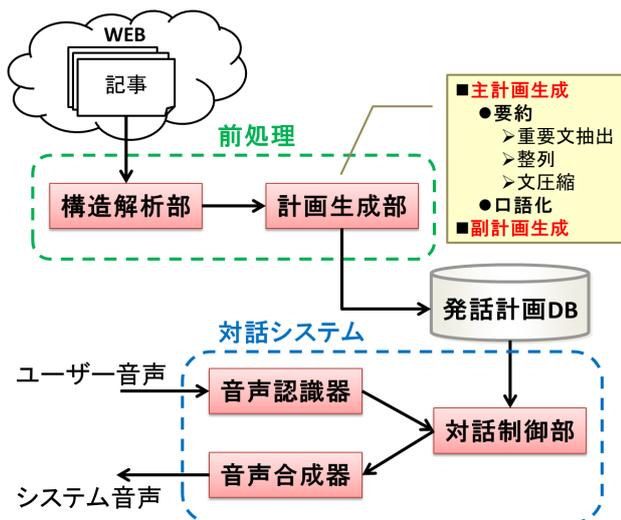


図 3: システム構成

### 3 想定する会話の例

本研究で想定している会話の例を図1に示す。会話の進行に伴い、ユーザーの意図性は大きく変化すると考えられる(図2)。これまでの多くの音声対話システムでは、ユーザーはシステムとのインタラクションに意図性の高い反応を示す必要があった。しかし、意図性の低い反応の中にも、ユーザーの理解状況を把握するのに有益な情報が含まれていると考えられ、本システムでは、言語的に明確な情報要求(e.g.「腰の痛みって?」)だけでなく、意図性の低い即応的な反応(e.g.「うん」や「え?」)も活用して、ユーザーの理解状況に合わせた情報提供を行う。

### 4 システムの要件とアーキテクチャ

快適な情報享受の実現に向けて次の観点から音声対話システムを設計する。

- (1) **簡潔明瞭性**：システムは、まとまりのある情報から要点をかいつまんで分かりやすく伝える。これにより、ユーザーは、興味がない内容を長々と聞かずに済む。まず、要約技術により要点を明らかにし、会話の主計画を立てる。そして、省いた情報から副計画を生成し、利用者の関心に応じて、副計画の内容を提供する。また、口語化処理を行い、話し言葉に近い柔らかい表現で情報を伝達する。
- (2) **インタラクティブ性**：システムは能動的な情報要求はもちろん、意図性の低い反応からも、ユーザーの意図をくみ取り、必要十分な情報を提供する。これにより、ユーザーは、欲しい情報を欲しいタイミングに、簡単な手段によって、手に入れられる。明示的な情報要求には、質問応答技術で対応し、非明示的な情報要求には、パラ言語情報などを利用し、テンポの良い会話を実現する。
- (3) **エンタテインメント性**：システムは、利用者に価値(知的刺激、安らぎ)を提供する。これにより、ユーザーは、本システムを気晴らしの手段として利用できる。会話自体がエンタテインメントとも言えるが、主観を交えてニュースの内容について議論できるようにしたり[10]、複数のキャラクタを選択できるようにしたり[11]することで付加価値を提供する。

システムの構成を図3に示す。まとまりのある情報の例題としてニュース記事を扱う。構造解析部では、収集したニュース記事に対して、形態素解析や係り受け解析などの基礎的な言語処理を行い、その結果を後段の処理で活用する。計画生成部では、主計画と副計画を生成する。主計画は、ニュースの要点に関する発話計画である。同じ話題について記述したニュース記事の集合を用いて複数文書要約を行い、その要約結果を口語化処理で話し言葉に書き換え、その結果から主計画を生成する。副計画は、主計画の情報を補うための発話計画である。副計画には、主計画で省かれた情報や予測される質問に対する回答、用語の定義などを組み込む。主計画と副計画からなる発話計画はデータベース上においておき、対話制御部でこれを参照しながら会話を進めていく。

### 5 主計画生成

主計画はニュースの要点に関する発話計画である。まず、同じ話題のニュース記事集合から要約に含めるべき文を抽出する(重要文抽出)。次に、抽出した文の提示順序を決定する(整列)。そして、重要文から不要な箇所を取り除き、文自体を短く要約する(文圧縮)。最後に、文を会話表現に書き換える(口語化)。

まとまりのある情報を伝える場合、伝える情報が多すぎるとユーザーは興味がない内容を最後まで黙って聞いていなければならない。一方で、伝える情報が少なすぎるとユーザーが新たな情報(興味を抱くかもしれない情報)に遭遇する機会を奪ってしまう可能性がある。そこで、見出し的な内容だけでなく、ユーザーの関心を引くような情報も加味した発話計画を立てる必要がある。このような観点で要約コーパスを設計し、そのコーパスを用いて、重要文抽出、整列、文圧縮の実験を行った[14]。

ニュース記事に代表される文書は書き言葉で記述されている。書き言葉は文字で情報を伝達することを目的に使われる言葉であり、そのまま音声で読み上げても、会話によるインタラクティブな情報伝達は実現できない。そのため、書き言葉として書かれた内容を話し言葉に書き換える必要がある。そこで、上記の要約コーパスを土台に口語化コーパスを設計し、話し言葉への書き換えを試みた[15]。

音声対話システムは、要約と口語化の処理を経てできた発話主計画に従って、ニュースの要点をインクリメンタルに(小出しにして)伝える。

### 6 副計画生成

副計画は主計画の情報を補うための発話計画である。テンポの良い会話を実現するためには、想定されるユーザーの反応に対して事前に発話計画を立てておく必要がある。副計画には、主計画で省かれた情報や予測される質問に対する回答、用語の定義などを組み込む。会話は主計画に沿って進行するが、ユーザーの情報要求に応じて、副計画に遷移し、必要十分な情報を提供する。

現在の音声対話システムでは、ユーザーの発話として質問や反復、感想などを想定している。このようなユーザーアクションに対して、発話意図を推定し、システムが返すべき応答を計画する。ユーザーアクションとその発話意図に対するシステムアクションの分類を表1に示す。

表 1: ユーザーアクションとシステムアクション

ユーザーアクション		発話意図	システムアクション
大分類	細分類		
相槌	弱い	無意識	継続
	強い	関心	肯定・感情共有・補足説明
「え」相当表現		確認	反復・真偽判定
		驚き	感情共有
		定義質問	定義型質問応答
		理由質問	why 型質問応答
反復	名詞	確認	反復・真偽判定
		驚き	感情共有
		自問	肯定・継続
		定義質問	定義型質問応答
	述語 フレーズ	確認	反復・真偽判定
		驚き	感情共有
		自問	肯定・継続
		理由質問	why 型質問応答
質問	明示的	質問	質問応答
	半明示的		
感想		感想	感情共有

発話意図は発話の表層表現だけでなく、文脈やパラ言語情報を用いて推定するが、意図性の低い反応による情報要求（非明示的な情報要求）は、その要求内容が明示的でないため、ユーザーの要求を一意に定めるのが難しい。例えば、「羽生結弦が／国際大会を／欠場するんだって」という主計画を述べた後に、ユーザーが「え」のような反応を示した場合、その発話意図が、羽生結弦が国際大会を欠場する理由に対する疑問なのか、驚きなのか、羽生結弦という人物に対する疑問なのか、それとも、全体を聞き逃したため、もう一度読み返して欲しいという意思表示なのか推定する必要がある。当然、ユーザーの発話意図を誤って推定してしまうことも想定されるが、このような誤解は会話のプロセス（否定したり、より厳密な質問に変えたり）で解消されることが期待できる。

現在、発話意図は規則に基づいて決定している。例えば、「え」という反応の直前の固有名詞が珍しい単語（IDF が小さい）ならば定義説明を行い、そうでなければ、反復して読み上げる。また、言語的な文脈情報以外に、パラ言語情報としてユーザー発話のピッチ（F0）変化の傾きを利用して「羽生結弦？」と「羽生結弦…」のような発話を区別している。

質問応答として、ファクトイド型と定義型、why 型、真偽判定型を扱う。質問は「メキシコオープンっていつ行われたの」という明示的な質問と「いつ」「誰」「何それ」という半明示的な質問に分けられ、半明示的な質問には文脈に基づいて回答する。

**定義型質問応答**：用語の定義を問う質問に回答する。記事内からは、同格や連体修飾、直接的な記述（e.g. 「NP とは～」）を利用して定義を獲得する。連体修飾からの定義獲得では、被修飾名詞の固有表現（e.g. 羽生結弦 ⇒ 人物）または Wikipedia における上位語情報（e.g. エボラ出血熱 ⇒ 急性ウイルス性感染症）を利用して、最近類と種差の考え方にに基づき、定義文を作成する。

記事外の定義情報として、Wikipedia の定義、EDR 辞書の定義、Web コーパスから獲得した定義 [10] を使用している。複数の定義文が候補となる場合もあるが、その場合は、定義候補集合内で出現する単語の頻度を計算し、頻出単語を多く含む定義文がランキング上位にくるようにしている。しかし、この方法では典型的な意味の定義文が選ばれるため、多義語には対処できない。そのため、記事の文脈に合わせた多義性の解消が今後の課題となる。

**述語項構造の含意矛盾認識**：以降で説明する真偽判定やファクトイド型質問応答、why 型質問応答が必要となる技術である。ここで、述語項構造が含意しているとは、次の二つの条件を満たすことである。

(1) 一方の述語が他方の述語を含意している：一方の述語が他方の述語を含意しており、両方の述語が同じ極性（肯定／否定）である（e.g. 敗北する ⇒ 負ける）。または、述語が反義関係にあり、一方が否定表現である（e.g. 負ける ⇒ 勝てない）。このとき、述語が含意関係にあると考える。

(2) 述語の格構造が類似している：全ての格に関して共通格の要素間で類似度を計算し、その平均が閾値以上であれば同一の述語項構造であると判断する。閾値は格要素の数が少ないほど高くなるように設定する。現在のシステムでは、格要素が一つするとき 0.8、それ以上するとき 0.7 としている。格要素の表層的な類似度は正規化したレーベンシュタイン距離（レーベンシュタイン距離を長い方の文字列の文字数で割った値）により計算する。ただし、辞書上で同義関係にある単語（e.g. 「米国」と「アメリカ」、「原発」と「原子力発電」）は類似度 1 とする。述語項構造が矛盾しているとは、次の二つの条件を満たすことである。

(1) 述語が反義関係にある：述語が反義関係にあり、両方の述語が同じ極性（肯定／否定）である（e.g. 負ける ⊥ 勝つ）。または、一方の述語が他方の述語を含意しており、一方が否定表現である（e.g. 敗北する ⊥ 負けない）。このとき、述語が反義関係にあると考える。

(2) 述語の格構造が類似している：含意の条件 2 と同じである。

ここで、動詞の含意知識として ALAGIN の『動詞含意関係データベース』を、格要素の同義知識として ALAGIN の『基本的意味関係の事例ベース』の異表記情報と同類語情報、および、NICT の『日本語 WordNet 同義対データベース』を使用している。反義知識は Web 上から人手で収集した反義ペアを『動詞含意関係データベース』を用いて拡張したものを使用している。

**真偽判定**：ユーザーの発話内容が記事の内容と一致しているかどうかを判定する。述語項構造の含意矛盾認識により実現する。例えば「錦織圭がダビド・フェレールに敗れた」という文があるとき、「錦織圭が負けたんだ」というユーザー発話に対しては「そうだよ」などの肯定表現で応じ、「錦織圭が勝ったの」というユーザー発話に対しては「違う。負けたんだよ。」のように、間違いを指摘した上で、正しい情報を与える。

**why 型質問応答**：理由を尋ねる質問に回答する。「ため」や「ので」などの特定のパターンに基づいて理由を獲得する。そして、理由の係り先の述語の述語項構造を用い

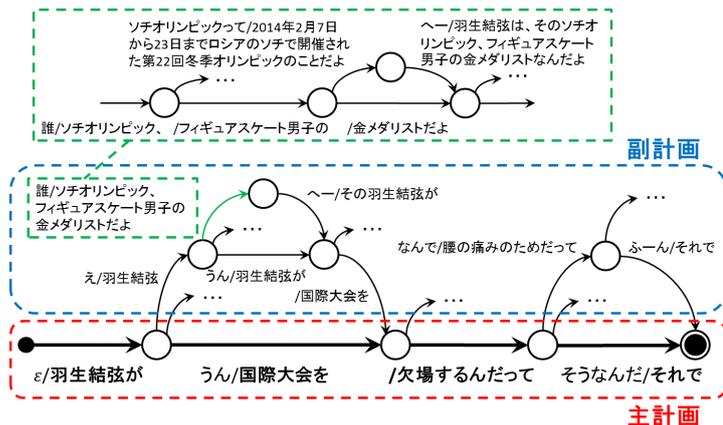


図 4: 発話計画に従った会話例

て、含意関係にある他の文の述語項構造に対して理由を付与する。これにより、文をまたいだ理由獲得が可能となる。明示的な質問に対しては記事全体から回答を探すが、「なんで」や「どうして」のような半明示的な質問に対しては直近の述語項構造に基づいて回答する。

ファクトイド型質問応答：単純な事実や出来事に関する質問に答える。明示的な質問に対しては記事全体から回答を探すが、「いつ」や「誰に」のような半明示的な質問に対しては直近の述語項構造に基づいて回答する。

述語項構造の含意認識を行い、同一の述語項構造がある場合、互いに不足する情報を補う。例えば、「メキシコオープンがアカプルコで行われた」という文と「メキシコオープンが2月28日に行われた」という文があるとき、前者に日付情報「2月28日に」を付与し、後者に場所情報「アカプルコで」を付与する。

現在は述語の項にくる名詞の固有表現や JUMAN のカテゴリ情報を利用して回答しているが、今後はより多様な質問に回答できるように拡張固有表現 [12] に対応させる。

## 7 音声対話システム

対話制御：主計画と副計画からなる発話計画に従った会話の例を図4に示す。 $\epsilon$  はユーザーの反応を見ずに状態が遷移することを表す。ユーザーが受動的な態度をとり続ける限り、システムは主計画に従った情報伝達を行う。ユーザーが情報要求を行った場合、その発話意図に合わせて副計画に遷移する。

ポーズの挿入：ユーザーは任意のタイミングで反応を示すことができるが、システムが淡々と話し続けていたのでは、インタラクションを取りにくい。そのため、適度にポーズを挿入する必要がある。現在のシステムでは、海木らのポーズ挿入規則 [13] を参考に、係り受け木において、当該文節の句が右枝分かれ、かつ、先行文節が左枝分かれである箇所、および、読点を含む箇所にポーズを入れている。

音声認識：{え、へー、いつ、誰、...} のような短い発話用の音声認識器と長い発話用の音声認識器を別々に用意する。adintool で発話区間検出を行い、短い発話認識には Julius を使用し、長い発話認識には Google の Web Speech API を使用している。また、同時にパラ言語認識も行うが、

現在は F0 変化の傾きのみを見ている。Julius の辞書には相槌や定型表現を事前に登録しておき、さらに、記事内の単語およびその派生形 (疑問形など) を自動で辞書に追加する。

音声合成：主に OpenJTalk を使用している (rospeex や DOCOMO の音声合成 API も指定できる)。発話計画に従った情報伝達を時間で管理するため、Julius の音素セグメンテーションキットを用いて、音声合成データから音素の読み上げ時間を算出している。しかし、この際、無音区間が無視されてしまうので、パワーを計算し、無音区間の時間を補完している。

## 8 おわりに

ニュース記事を例題としてまとまりのある情報を利用者が快適に享受するための仕組みを考え、音声対話システム上での実現を試みた。システムは、主計画 (要点) に沿った説明を行い、利用者の反応に応じて副計画 (補足) の情報を提示する。利用者は、基本的には受け身で利用できるが、理解や興味の状況に応じて、意識的・無意識的にシステムとインタラクトすることで、必要十分な情報を得ることができる。今後は、本システムの有用性を確認するための評価実験を計画し実施する。

### 参考文献

- [1] D.O.Case: "Looking for Information", A Survey of Research on Information Seeking, Needs, and Behavior, Second Edition, Academic Press, 2007.
- [2] K.E.Fisher, S.Erdelez, and L.E.F.McKechnie: "Theories of Information Behavior", Information Today, 2005.
- [3] S.Erdelez: "Information encountering", in [3], pp.179-184, 2005.
- [4] S.Seneff and J.Polifroni: "Dialogue management in the Mercury flight reservation system", in Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems, Vol.3, pp.1116, 2000.
- [5] 駒谷和範, 上野晋一, 河原達也, 奥乃博: "ユーザモデルを導入したバス運行情報案内システムの実験的評価", 情報処理学会研究報告, SLP-47-12, pp.59-64, 2003.
- [6] 杉山聡, 堂坂浩二, 川端豪: "音声対話によるテキスト内容の伝達方法", 情処学論, Vol.41, No.6, pp.1883-1894, 2000.
- [7] 翠輝久, 河原達也, 正司哲朗, 美濃導彦: "質問応答・情報推薦機能を備えた音声による情報案内システム", 情処学論, Vol.48, No.12, pp.36023611, 2007.
- [8] Y.C.Pan, H.Y.Lee, L.S.Lee: "Interactive spoken document retrieval with suggested key terms ranked by a Markov decision process", IEEE Transactions on Audio, Speech, and Language Processing, Vol.20, No.2, pp.632645, 2012.
- [9] 吉野幸一郎, 河原達也: "ユーザの焦点を用いた POMDP による音声情報案内システム", 人工知能学会言語・音声理解と対話処理研究会, 2014-SLUD72, 2014.
- [10] 高津弘明, 小林哲則: "多人数会話における議論支援のための状態把握", 言語処理学会第 22 回年次大会発表論文集, 2016.
- [11] 高津弘明, 小林哲則: "対話エージェントのための性格モデル", 言語処理学会第 21 回年次大会発表論文集, pp.191-194, 2015.
- [12] S.Sekine, K.Sudo, and C.Nobata: "Extended Named Entity Hierarchy", in Proceedings of the Third International Conference on Language Resources and Evaluation, 2002.
- [13] 海木延佳, 句坂芳典: "局所的な句構造によるポーズ挿入規則化の検討", 信学論 (D-II), Vol.J79-D-II, No.9, pp.14551463, 1996.
- [14] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: "会話によるニュース記事伝達のための情報選択", 言語処理学会第 22 回年次大会発表論文集, 2016.
- [15] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: "会話によるニュース記事伝達のための口語化における述語の書き換え", 言語処理学会第 22 回年次大会発表論文集, 2016.
- [16] 林良彦, 藤江真也, 福岡維新, 高津弘明, 小林哲則: "情報アクセスにおける受動性と能動性:音声対話によるニュース記事アクセス", 人工知能学会第 11 回インタラクティブ情報アクセスと可視化マイニング研究会, pp.25-32, 2015.
- [17] 藤江真也, 福岡維新, 麥田愛純, 高津弘明, 林良彦, 小林哲則: "効率的な情報伝達を指向した音声対話システムの提案", 人工知能学会第 74 回言語・音声理解と対話処理研究会, SIG-SLUD-B501-02, 2015.