

## 音声認識/文章解析の誤り訂正機能を持つ対話処理

清水 友裕 野村 浩郷

九州工業大学情報工学部

〒820-8502 飯塚市川津 680-4

nomura@ai.kyutech.ac.jp

<http://www.dumbo.ai.kyutech.ac.jp/nomura/>

**概要** : 自然言語対話において, 音声認識や文章解析での誤りによって適切な応答が困難になるという問題がある. 本稿では, そのような誤りを訂正する対話を対話全体の一部として自然な形で埋めこみ, 誤り訂正をしながら対話を適切に進行させる対話処理について述べる. 対話のドメインとしては, 便宜的に, パソコン技術サポートを行うコールセンターでの質問応答を取上げ, 既存の大量な質問応答データから取得したドメイン知識などを適用して, 誤り訂正を行う対話処理について考察する.

## Enhanced Dialogue System Equipping Error-correcting Function on Voice Misunderstanding and Sentence Incorrect Analysis

Tomohiro Shimizu Hirosato Nomura

Kyusyu Institute of Technology

Iizuka 820-8502 Japan

nomura@ai.kyutech.ac.jp

**Abstract** : In a natural language dialog, there is a problem that a suitable response becomes difficult by the mistake in speech recognition and/or the error in text analysis. In this paper, the dialog which corrects such errors is incorporated in a form natural as a part of whole dialog, and the dialog processing which advances a dialog appropriately carrying out an error correction is discussed. As a domain of the dialog, the question-answering at the call-center which performs personal computer technical support is taken up, for our convenience. The domain knowledge is acquired from large amount of question-answering data, and is applied to the dialog processing which performs an error correction.

## 1 はじめに

今日、インターネットや電子メールの利用により、パソコンが家庭の中に急速に普及してきている。パソコンを操作する際、ユーザはキーボードやマウスを使ってパソコンと対話をするが、人間同士の対話とは違い、ユーザがキーボードでコマンドを入力したり、マウスでクリックするなどしてパソコンに要求を伝え、パソコンはその要求にしたがって処理をし、その結果を表示するというように、その対話方法は何も知らない初心者にとってはまだまだ困難なものと言える。

パソコンを操作する上で、こうした基本的な技術や知識すら必要とせず、誰でも容易かつ手軽にパソコンと対話するには、やはり人間が用いる自然言語でコンピュータと対話でき、さらに、入力インターフェースとしてキーボードなどではなく音声認識を用いることが望ましい。

音声による自然言語対話を行うためには、入力された音声を音声認識部で文字情報に変換し、入力解析部で文章解析し、問題解決部で応答情報収集や対話制御処理などにより応答内容を決定し、発話生成部で生成した応答文を音声出力部を用いて出力するといった処理が必要である。このとき、音声認識部では認識誤り、入力解析部では文章解析誤りが起こる可能性があり、これらの誤りによって問題解決に失敗してしまい、適切な応答が困難になるという問題がある。そのため、音声対話システムを実現するにあたっては、それぞれの誤りに対する誤り訂正機能が必要である。

そこで、本研究は自然言語対話において、音声認識や文章解析での誤りに対して、それらを訂正する対話を自然な形の対話の一部に組み込み、入力の誤りを訂正しながら対話を適切に進行させる対話処理の実現を目指す。

本稿では、対話のドメインとして、パソコン技術サポートを行うコールセンターでの質問応答を取上げ、メールコールセンターでの約3年間の実務により収集された約35,000件の質問および回答のメールデータから取得したドメイン知識などを適用して、誤り訂正を対話で行う対話処理について考察する。なお、音声認識エンジンとしては、市販のPC用音声認識ソフトウェアを使用する。対話は画像なども使うマルチモーダル対話であるが、本稿では音声による対話の部分に限定して述べる。

## 2 自然言語対話における問題点

音声による自然言語対話を実現するには次のような問題がある。

### 2.1 音声認識誤り

キーボードを使った文字入力では、文字を打ち間違えたり文法的な間違いがあったときには、ユーザが入力前にある程度訂正して入力することができるが、音声入力では思いつくままに発話したものがそのまま認識され、入力されてしまうので、認識誤りや文法誤りなどを含む可能性が高くなる。音声認識に誤りがあると、次の文章解析でも誤った入力に基づいた解析で誤りが拡大してしまい、問題解決に大きな障害となる。

以下に音声認識誤りの例を示す。

**発話文：** 電源が入りません  
**認識結果：** 電源がはよません

音声認識の誤り方には以下のようなものがある。

- 音韻認識の誤りによる誤認識  
ユーザの発音やマイクの不調が原因で音の一部ないしは全部がなかった場合。
- 音素列に混入した雑音音素による誤認識  
ユーザの意図していない発声や周囲の雑音により誤認識が起きた場合。
- 似た音素の誤認識  
母音が同じ音で、音素が似ているため誤認識が起きた場合。
- 音素列の不適切な区切りによる誤認識  
音声を単語に変換する時に単語の境界を間違ってしまった場合。例えば、1つの語の発声の途中に空白が入った場合。

本研究では市販の音声認識ソフトを使っているため、音韻レベルの問題には関心しない。

さらに、音声認識には

- 同音異義語の変換の誤り

もあり、これも自然言語対話の実現にあたって対処すべき問題となる。

## 2.2 文章解析誤り

文章解析としては、形態素解析、構文解析、意味解析、文脈解析、談話解析などがあげられる。現在、形態素解析、構文解析はかなりの技術が確立しており、日本語の構文解析としては係り受け解析がよく使われている。形態素解析器ではJUMAN, ChaSen, MeCab, 係り受け解析器ではKNP, CaboChaなどが広く利用されている。これらはかなりの精度で処理するが、誤りも少なくない。

質問応答といったような意味のある対話をしようとする時、形態素解析、構文解析に続いて、意味解析なども行わなければならない。形態素解析で誤ると構文解析に影響し、構文解析で誤ると意味解析に影響するため、これらの精度がかなり高いとは言え、現実起こる誤りを無視することはできない。

以下に係り受け解析での誤りの例を示す。

- CaboChaでの解析誤りの例

入力文：メールの最初の設定法を知らないの  
で教えていただきたい

解析結果：

メールの-D  
最初の-D  
設定法を-D  
知らないの-D  
教えていただきたい

この例では、「メールの」は「最初の」に係っているが、文の意味を考えると「メールの」は「設定法を」に係るはずである。

係り受け解析の誤りは、文が以下の要素を持つとき起こりやすい。

- 連体修飾語
- 従属節
- 並列構造

音声認識と文章解析で共通して問題なのが、それぞれの精度を上げたとしてもそれ単体では100%の精度は得られないであろうという点である。なぜなら、自然言語は本来曖昧さを含むものであり、意味や文脈を考慮しなければ、音声認識や形態素解析、構文解析時には決定できない部分が残るからである。また、人間が自由に入力できる以上、入力文自体が文

法的に間違っている可能性もある。よって、自然言語対話には、対話をしながらこれらの誤りを訂正する仕組みが必要となる。

## 3 対話の制御

話し手が聞き手に対してどのような発話で情報を伝達すべきかという原則として、一般的に以下に示すGriceの公準が知られている。

1. 量の公準  
過不足のない情報を伝える。
2. 質の公準  
根拠のある真実や真であると思うことを告げる。
3. 関係の公準  
話し手と聞き手のお互いの関連した事柄を話す。
4. 様態の公準  
明瞭に簡潔に順序立てて話す。

これを誤り訂正の対話での発話に当てはめ、次のよう考える。

1. 量の公準  
音声認識や文章解析の誤りを訂正するために必要であると判断する情報を過不足なく伝える。
2. 質の公準  
入力文について、音声認識や文章解析の誤りではない、または誤りであると判断できた部分を告げる。
3. 関係の公準  
入力文に対する音声認識、文章解析の誤りについて話す。
4. 様態の公準  
入力文のどの部分に対して、誤りと判断しているかがはっきり伝わるように話す。

以下に示す誤り訂正対話では、これを満たすように発話を作成する。例えば、「電源がはよません」という音声認識の誤りに対して、入力文を全部用いて「電源がはよませんと言いましたか?」「電源がはよまないのですか?」等の発話は量の公準や質の公準を満たさない。この場合、正しいと判断される「電源が」を用いて「電源がどうしました?」「電源が何ですか?」等の発話が望ましい。

## 4 音声認識誤りの訂正

### 4.1 語の共起

約 16,500 件のメールデータから質問文を抽出し、46,224 文について、形態素解析器 MeCab を用いて形態素解析を行い、品詞が名詞、動詞、形容詞のもので非自立ではなく、かつ数でない語の同文中での他の語との共起度  $C$  を調べたところ、13,942 語についての共起度が得られた。

1 文中において、ある語  $a$  の他の語  $b$  との共起度  $C(a, b)$  は、全質問文中での共起頻度  $cf(a, b)$ 、語  $a$  と共起した語の数  $cn(a)$  を用いて、次式のように定義する。

$$C(a, b) = cf(a, b) / cn(a)$$

この値は語  $a$  の他の語  $b$  との同じ文中での出現しやすさを表しているのので、入力文中の各語  $x$  について、文中の他の語  $w_i$  との共起度の和

$$C(x) = \sum_i^n C(x, w_i)$$

を求めると、 $C(x)$  が低いものは、音声認識誤りによって発話とは違う語が認識された可能性が高く、逆に  $C(x)$  が高いものは、正確に認識できている可能性が高いと推測できる。

### 4.2 訂正手法

音声認識の誤りを訂正する対話は、前述の語の共起を使って誤りがあるかどうかの判定をし、ある場合は対話の制御に基づく応答として、言い直しを要求するという手法をとる。

まず、認識された質問文を形態素解析し、品詞が名詞、動詞、形容詞のもので非自立ではなく、かつ数でない語を取り出す。取り出した各語について他の語との共起度の和  $C(x)$  を前述の方法で計算し、値が閾値より低い語を認識誤りの語と判定する。閾値は本論文では便宜的に 0.01 とする。例えば「パソコンの動作がとても遅い」という入力文に対しては、形態素解析の結果が

パソコン 名詞, 一般, \*\*\*\*, パソコン, パソコン, パソコン  
の 助詞, 連体化, \*\*\*\*, の, ノ, ノ  
動作 名詞, 名変接続, \*\*\*\*, 動作, ドウサ, ドーサ  
が 助詞, 格助詞, 一般, \*\*\*\*, が, ガ, ガ  
とても 副詞, 助詞類接続, \*\*\*\*, とても, トテモ, トテモ  
遅い 形容詞, 自立, \*\*, 形容詞・アウオ段, 基本形, 遅い, オソイ, オソイ

となり、品詞が名詞、動詞、形容詞で非自立ではなく、かつ数でない「パソコン」「動作」「遅い」が取り出される。「パソコン」について他の 2 語との共起度の和を計算すると、

$$\begin{aligned} C(\text{パソコン}) &= C(\text{パソコン}, \text{動作}) + C(\text{パソコン}, \text{遅い}) \\ &= 0.139954790147049 \end{aligned}$$

となり、同様に「動作」「遅い」についても計算すると、

$$C(\text{動作}) = 0.0892583680999173$$

$$C(\text{遅い}) = 0.0764621645888845$$

という結果になる。この場合どの値も閾値以上なので、正しく認識された語と判断できる。

誤りと判定された語がある場合は、言い直し要求文を作成する。言い直しでは、余計な音声認識誤りを避けるために、正しく認識できた語の再入力を排し、誤り部分を含み、できるだけ前の入力より短い文であるほうがよい。よって作成する要求文は、質問のどこで認識誤りが起こったかがはっきり分かるものである必要がある。そこで、共起度を計算した語のうち、誤りと判定された語の 1 つ前の語までの入力文を使い、その後ろに「何ですか？」等と付け加えて言い直し要求文を作成する。また、誤りと判定された語の 1 つ前の語の品詞が名詞の場合、その直後に助詞がある時は、言い直し要求文が自然な形になるようにその語も付け加える。例えば「検索ワードの履歴を削除する方法を教えてください」という発話が「検索ワードの履歴を作物方法を教えてください」と誤認識された場合、

検索	名詞, 名変接続, ****, 検索, ケンサク, ケンサク
ワード	名詞, 一般, ****, ワード, ワード, ワード
の	助詞, 連体化, ****, の, ノ, ノ
履歴	名詞, 一般, ****, 履歴, リレキ, リレキ
を	助詞, 格助詞, 一般, **, を, ヲ, ヲ
作物	名詞, 一般, ****, 作物, サクモツ, サクモツ
方法	名詞, 一般, ****, 方法, ホウホウ, ホーホー
を	助詞, 格助詞, 一般, **, を, ヲ, ヲ
教え	動詞, 自立, **, 一段, 連用形, 教える, オシエ, オシエ
て	助詞, 接続助詞, ****, て, テ, テ
下さい	動詞, 非自立, **, 五段・ラ行特殊, 命令 i, 下さる, クダサイ, クダサイ

と形態素解析され、 $C(\text{作物}) = 0.0$  になり名詞「作物」が誤りと判定されるので、「作物」の 1 つ前の語までの入力文「検索ワードの履歴」と名詞「履歴」の直後の助詞「を」を使って、

「検索ワードの履歴を何ですか？」

という言い直し要求文が作成される。この言い直し要求文には、例えば「検索ワードの履歴がどうしま

したか？」等、対話の制御に従って決まる色々な表現がある。

この言い直し要求に従って言い直された入力文を、元の入力文と比較し、誤りと判定されていた語が異なっていた場合は置換して対話を次へ進める。同じ場合はその語だけ使って「ですか？」または「と言いましたか？」と付け加えて Yes/No で返事ができる確認文を作成する。返事が Yes であればそのまま対話を次へ進め、No であれば言い直し要求文からやり直す。

## 5 文章解析誤りの訂正

### 5.1 語の係り受け

質問文 46,224 文に対して、係り受け解析器 CaboCha を用いて係り受け解析を行った。解析結果から、品詞が名詞、動詞、形容詞のもので非自立ではなく、かつ数でない語について、ある語 a が語 b に係っているとき、語 a の語 b との係りやすさを係り受け度  $D(a, b)$  として調べた。係り受け度  $D(a, b)$  は、全質問文中で語 a と語 b が係り受け関係として出現した頻度  $df(a, b)$ 、a と係り受け関係になった語の数  $dn(a)$  を用いて、次式のように定義する。

$$D(a, b) = df(a, b) / dn(a)$$

このとき語 a の品詞が名詞で、その直後に助詞 c があるときは、語 b に助詞 c を加えて考え、助詞が違う係り受け関係は別のものとして扱った。

$D(a, b)$  は語 a (または語 a と助詞 c) の語 b との係りやすさを表しているのだから、入力文を係り受け解析したときに、 $D(a, b)$  が低いものは、係り受け解析が誤っている可能性があり、逆に  $D(a, b)$  が高いものは、正確に係り受け解析ができていない可能性が高いと推測できる。

前述のように、係り受け解析は誤っている可能性があるのだから、この値を過度に信用することはできないが、大量の文を解析して頻度を調べているのだから、係り受け解析の精度から言って、誤りが大量に出てくることはないかと期待して、どの語を係り受け解析誤りの訂正対象にするかの判断材料として使う。

### 5.2 訂正手法

文章解析誤りの訂正は、本論文では係り受け解析の訂正に絞って考える。前述の語の係り受け度を使っ

て、入力文の係り受け解析結果で誤りの可能性がある係り受け関係を探し、見つかった場合は係り受けを確認するという手法をとる。

まず、入力文を係り受け解析し、入力文の各文節に対して品詞が名詞、動詞、形容詞のもので非自立、数でない語を取り出す。このとき、文節内で名詞が連続しているときは名詞として 1 つにまとめ、名詞の後に動詞「する」があるものは合わせて動詞とする。また、名詞の後に助詞があるものはその助詞も取り出しておく。さらにその文節が係っている文節内でも同様に名詞、動詞、形容詞の語を取り出し、係り受けの組を作る。各係り受けの組について、前述の係り受け度  $D(a, b)$  を求め、さらに語 a の文中に出てくる他の語  $w_i$  との  $D(a, w_i)$  も計算する。もし、 $D(a, b)$  よりも大きい  $D(a, w_i)$  が存在する際は語 a と語 b の係り受けを確認するような発話を行う。

例えば、「メールの最初の設定が分からない」という入力文に対しては、係り受け解析の結果が

メールの-D  
最初の-D  
設定が-D  
分からない

となり、 $D(\text{メールの, 最初})$ 、 $D(\text{最初の, 設定})$ 、 $D(\text{設定が, 分かる})$  が計算される。

「メールの」に対しては  $D(\text{メールの, 設定})$ 、 $D(\text{メールの, 分かる})$  が計算され、

$$D(\text{メールの, 最初}) = 0.0$$

$$D(\text{メールの, 設定}) = 0.0177993527508091$$

$$D(\text{メールの, 分かる}) = 0.0$$

となり、「メールの」が「設定」に係るほうが係り受け度が大きいことから「メールの最初ですか？」という Yes/No で返事ができる確認文を作成する。返事が Yes であればそのまま次の対話へ、No の場合は「メールの設定です」といった訂正を含む返事ならば「メールの」の係り先を「設定」に直して次の対話へ、訂正を含まない返事ならばメールの係り受け度の高かった「設定」を用い、「メールの設定ですか？」という確認文を作成する。

## 6 実験

本稿で述べた手法を用いた対話例を以下に示す。

- 音声認識誤りの訂正対話例

ユーザ：文字を入力してる途中いきなり文字がへんてこなところに飛びます  
（「文字を入力してる途中いきなり文字がへんでもなところに飛びます」と誤認識）  
システム：文字を入力してる途中いきなり文字が何ですか？

ユーザ：変なところに飛びます  
（「文字を入力してる途中いきなり文字が変なところに飛びます」と入力文を訂正）

- 文章解析誤りの訂正対話例

ユーザ：子供がインストールしてあったアプリケーションを削除してしまった  
（解析で「子供が」が「インストールする」に係る）  
システム：子供がインストールしてあったのですか？  
ユーザ：子供が削除してしまったんです  
（「子供が」の係り先を「削除する」に訂正）

## 7 おわりに

本稿では、自然言語対話において、音声認識や文章解析での誤りがある場合でも、それらを訂正する対話を対話全体の一部として自然な形で組み込み、誤り訂正をしながら対話を適切に進行させる対話処理について考察した。

今回は、対話による誤り訂正の基本的な機能を持つシステムを作成して実験を行ったところ、有効に働くことが確認できた。システムとしての実用に耐える対話をするには、より複雑な機能を持つ必要があるので、それを追加することで、より複雑な誤り訂正をできるよう発展させるのがよいと思われる。よって今後は、正確な訂正、適切な対話を実現すべく、これらの機能を詳細化し、それに基づくシステムを作成する予定である。

また、音声認識や文章解析の誤り訂正に関しては、未知語にどう対応するかの問題もあり、本稿ではそれに対する処理は述べなかったが、これについては別の機会に述べる予定である。

## 参考文献

- [1] 野村 浩郷. 自然言語処理の基礎技術, 電子情報通信学会, 1988.
- [2] 岸川 寿. 話台辞書と音声対話による音声認識誤りの修正, 平成 15 年度九州工業大学卒業論文, 2003.
- [3] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡一馬, 浅原 正幸. 形態素解析システム『茶筌』 version 2.3.3 使用説明書, 奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学講座, 2003.
- [4] 工藤 拓, 松本 裕治. チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp.1834-1842, 2002.
- [5] 黒橋 禎夫, 河原 大輔. 日本語形態素解析システム JUMAN version 5.1, 東京大学大学院情報理工学系研究科, 2005.
- [6] 黒橋 禎夫, 河原 大輔. 日本語構文解析システム KNP version 2.0, 東京大学大学院情報理工学系研究科, 2005.
- [7] 浜辺 良二, 駒谷 和範, 尾形 哲也, 奥乃 博. 音声対話システムにおける音韻の類似表現の混同を防ぐための確認の自動生成, 情報処理学会研究報告, 2005-NL-167, Vol.2005, No.50, pp.89-94, 2005.
- [8] 石川 開, 隅田 英一郎. テキストデータを使った音声認識誤りの訂正, 自然言語処理, Vol.7, No.4, pp.205-227, 2000.
- [9] 河原 達也, 荒木 雅弘. 知の科学 音声対話システム, 人工知能学会編, オーム社, 2006.
- [10] 緒方 賢史. 旅行申込対話モデルとそれに基づく旅行案内システムの研究, 平成 8 年度九州工業大学修士論文, 1996.
- [11] Paul Grice. "Logic and conversation". In *Syntax and Semantics*, 3: Speech Acts, ed. P. Cole & J. Morgan. New York: Academic Press, pp 41-58, 1975.