自然言語処理 106－8
（1995. 3. 10）

# 不完全文章処理としての音声理解

横田 将生　小田 まり子†　小田 誠雄
福岡工業大学　言語情報工学研究所
†久留米工業大学工学部　電子情報工学科

音声認識の理想は、話者の発話を文章として一字一句忠実に再現することである。しかしながら、実際には、そのことは極めて困難である。我々が構築しようとしている音声理解システムは、内容的に妥当であれば、発話者の表現にはこだわらない出力を行わせるものである。このような処理は、もはや、音声の波動としての物理的な特徴に関する処理範囲を超えた自然言語理解処理、特に、概念処理の領域に属するものとなる。そして、不完全な音素列あるいは文字列として知覚された音声情報の理解は、自然言語理解研究における重要な課題一つである省略語概念推定処理とほぼ等価となる。

# Speech Understanding as Incomplete Text Processing

Masao YOKOTA　Mariko ODA†　Seio ODA
Language and Information Laboratory, Fukuoka Institute of Technology
†Department of Information Science and Electronics Engineeirng,
Kurume Institute of Technology

The ideal goal of speech recognition is to reconstruct literally exact utterances. Actually, however, such a thing is quite difficult and almost impossible. We intend to construct such a speech understanding system as can infer the conceptual information which the speaker would transmit. The processing for the purpose belongs no longer to wave signal processing but to natural language understanding, especially, to conceptual processing. And moreover, understanding incompletely percepted speech is nearly equal to estimating the concepts of the words omitted in texts.

## 1．まえがき

音声認識の理想は、話者の発話を文章として一字一句忠実に再現することである。しかしながら、実際には、そのことは極めて困難である。我々が構築しようとしている音声理解システムは、内容的に妥当であれば、発話者の表現にはこだわらない出力を行わせるものである。このような処理は、もはや、音声の波動としての物理的な特徴に関する処理範囲を超えた自然言語理解処理、特に、概念処理の領域に属するものとなる。そして、不完全な音素列あるいは文字列として知覚された音声情報の理解は、自然言語理解研究における重要な課題一つである省略語概念推定処理とほぼ等価となる。

以下では、まず、音声情報伝達モデルおよび概念処理主導型音声理解モデルの提示を行い、次に、概念処理の核となる常識および状況知識(合わせて、背景知識と呼ぶ)および適切な理解結果を得るための工夫、最後に、今後の課題について報告する。

## 2．音声による情報伝達のモデル

人間$M_1$(話者)が人間$M_2$(聴者)に概念的内容(または情報)cを音声言語表現(音素列)rによって伝達する場合を想定してみよう。$M_1$がcの音声言語表現として発話しようとするrは、$M_1$自身の発声過程の誤り、音声伝搬路におけるノイズの混入、あるいは、

$M_2$自身の聴取過程の誤りにより複数の音声言語表現の集合$R_2$として知覚される。更に、$R_2$の各要素は$M_2$により意味解釈され概念的内容の集合$C_2$として曖昧に情報伝達されたことになる。次の式(1)－(3)は、以上の事柄を形式化したものである。

$$r \in \Phi_1(c) = R_1 = \{r_{11}, \ldots, r_{1l}\} \quad (1)$$
$$R_2 = \Delta_{12}(r) = \{r_{21}, \ldots, r_{2m}\} \quad (2)$$
$$C_2 = \bigcup \Phi_2^{-1}(r_{2i}) = \{c_{21}, \ldots, c_{2n}\} \quad (3)$$

ただし、

$\Phi_i$：$M_i$における概念的内容を音声言語表現に変換する過程(言語化過程)

$\Phi_i^{-1}$：$M_i$における音声言語表現を概念的内容に変換する過程(意味解釈過程)

$\Delta_{ij}$：$M_i$および$M_j$間での音声言語表現を歪ませる要因を関数で表現したもの

## 3.概念処理主導型音声理解のモデル

音声認識の理想は、上記式(2)の$R_2$の中から、rに等しい$r_{2i}$を選択することである。しかしながら、実際には、そのことは極めて困難であり、聴者$M_2$は、$C_2$の中から、常識(common sense)や状況(situation)に関する知識(状況知識と呼ぶ)に適合したものc'を選択し、それを話者$M_1$が伝達しようとした情報cとして推定することになる。この場合、c'とc、また、c'を担っている表現$r_{2i}$とrとは必ずしも等しくない。すなわち、聴者$M_2$は、内容的に妥当であれば、$r_{2i} \neq r$であるような$r_{2i}$を選択しうる。このような処理は、もはや、音声の波動としての物理的な特徴に関する処理範囲を超えた自然言語理解処理、特に、概念処理の領域に属するものとなる。我々が構築しようとしている音声理解システム(図1参照)はそのような処理を実現するものである。このようなシステムは、仕事(task)の一つとして、内容的に妥当であれば、発話者の表現にはこだわらない書き取り(dictation)結果を出力し、例えば、ある方言での会話"「どさ。」「湯さ。」"に対して"「(あなたは)何処へ(行くのですか)。」「(私は)風呂へ(行きます)。」"というように理解結果を表出することもありうる。

## 4．常識および状況知識

常識は、長期記憶に相当するが、状況知識は、短期または中期記憶に相当する。すなわち、比較的、常識は時間的に安定した静的なものであり、状況知識は逆に時変的(time-variant)で動的なものである。現在のところ、常識および状況知識としては以下のようなものを考えている。

### 4.1 常識の範疇

人間が長期記憶として共通に持っているものが常識と呼ばれる。しかしながら、各人の持つ「長期記憶としての知識」自体が明らかでないため、その共通部分など論じようのないのが実情である。したがって、ここでは、我々が自然言語理解処理のために提唱している心像意味論に準拠した長期記憶の範疇化を提示するにとどめる。心像意味論では、長期記憶をその対象世界あるいは議論世界(world of discourse)という観点から以下のように分類している。言語すなわち記号とそれが担う概念に関係し、主に、意味論(semantics)の取扱い範囲である。

(1) 世界に依存する長期記憶

日常世界や特殊世界に言及したり、影響を及ぼしたりする時に必要な知識

(2) 世界に依存しない長期記憶

主に論理的思考に必要な公理や推論規則などの知識

純粋な音声理解の立場で必要なのは、言語に関する知識が大部分である。

### 4.2 状況知識の範疇

音声理解(自然言語理解)の立場から、状況を、その発話がなされるに至った脈絡(context)と見なし、情報伝達媒体という観点で分類すると以下のようになる。

(1) 言語的脈絡(linguistic context)

人間(聴者$M_2$)が、ある時点(発話の時点)までに言語表現を介して収集した全ての情報

(2) 非言語的脈絡(non-linguistic context)

人間(聴者$M_2$)が、ある時点(発話の時点)までに言語以外の媒体を介して収集した全ての情報($M_1$の動作や表情など)

純粋な音声理解の立場では、状況は「言語的脈絡」だけに限られるが、話者が言語的に指示した環境などから聴者が感覚器官を介して得た情報は「非言語的脈絡」を発生させる可能性がある。このような事情から、主に語用論(pragmatics)の取扱い範囲である。

## 5．概念処理

聴者$M_2$が話者$M_1$からの伝達情報cの推定、すなわち、c'の選定を行う場合には、常識や状況知識を利用する。$M_2$のこのような行為に相当する音声理解

システムの処理過程は、概略、以下のようである。ただし、談話(discourse)とは、発話(utterance)rの順序集合と考えている。

[STEP1] 短期記憶にあるR₂の各要素の解釈を言語知識(長期記憶の一部で、形態素、構文、および意味に関する情報)を利用して行い、解釈結果の集合C₂を短期記憶として得る。

[STEP2] C₂の各要素に局所的選択公準(長期記憶の一部、後述)を適用して、不適切な解釈結果を除去した集合L₂(局所的理解結果の集合と呼ぶ)を短期記憶として得る。

[STEP3] L₂の各要素、および、その時点までの談話理解結果の集合G₂(中期記憶の一部)に大局的選択公準(長期記憶の一部、後述)を適用して、不適切な理解結果を除去し、新たな談話理解結果の集合G₂を中期記憶として得る。

[STEP4] 話者M₁の全発話(一つのまとまった談話)が終了した時点で、得られたG₂を構成している各発話rの理解結果をc'とする。

[STEP5] 理解結果c'が複数個存在する場合は、適切度評価関数(後述)を適用して、それらに優先順位を付ける。

[STEP6] 理解結果c'を優先順位付きで文章として表出する。

なお、STEP1-3における処理に関しては、図2を参考にされたい。

## 6. 選択公準

適切な発話の概念内容を推定するために用いる規則を選択公準(Postulation for Selection)と呼ぶ。これらは適用する範囲の大きさにより、局所的選択公準(PLS: Postulation for Local Selection)および大局的選択公準(PGS: Postulation for Global Selection)に分類する。前者は、単一の発話に、後者は、それまでになされた発話の順序集合に対して適用され、概念内容が不適切なもの(主に、冗長、矛盾などを引き起こすもの)を除去あるいは優先順位を低くすることにより妥当な発話および談話内容を絞り込むものである。これらの公準は、自然言語理解システムIMAGESを作成した経験を基に長期記憶の一部として設定しており、一種のメタ公理として用いられる。

### 6.1 局所的選択公準

現時点において、以下のような公準および適用例を考えている。

[PLS1] "無意味な発話は存在しない。"

(例1) 100gの土地を購入した。

[PLS2] "自己矛盾した発話は存在しない。"

(例2) 赤い花は青い。

[PLS3] "冗長な発話は存在しない。"

(例3)馬から落ちて、落馬した。

### 6.2 大局的選択公準

現時点において、以下のような公準および適用例を考えている。

[PGS1] "相互に無意味な発話を含む談話は存在しない。"

(例4)虹を見た。その重さは100gであった。

[PGS2] "相互に矛盾した発話を含む談話は存在しない。"

(例5)赤い花が咲いている。その色は、青である。

[PGS3] "相互に冗長な発話を含む談話は存在しない。"

(例6)馬から男が落ちた。その男は落馬した。

## 7. 言語理解処理

図1のシステムにおいて、音声認識部(Speech recognition)で不完全に認識された文章が、言語理解部(Language Understanding)に渡された場合の処理の概略を以下に説明する。

### 7.1 省略概念の推定

音声理解処理は、離散的に認識された単語概念間の欠落した単語(句)概念を陽に記述することが目的となり、そのためにシステムは常識および状況知識(これらを合わせて、背景知識と呼ぶ)を利用する。その場合、全ての語の意味記述が与えられているとの前提に立っており、離散的に認識された単語の意味構造がそれを含む文脈の意味構造に矛盾なく統合されたときにその理解処理は成功したことになる。このことは、式(4)における推理記号(⊢)の右辺(すなわち、欠落語を補完した構造の意味解釈)が導出されるような処理を行うことを意味する。ただし、この式で、Bは背景知識である。表層的には式(5)に示すように、離散的に単語認識された音声情報Pの欠落箇所$x_i$へ挿入される語句$p_i$の集合(代入)$\theta$を求めることになる。

$$I(P[x_1,\ldots,x_n]) \wedge B \vdash I(P[p_1,\ldots,p_n]) \quad (4)$$

$$I(P) \wedge B \vdash I(P\theta), \quad \theta = \{x_1/p_1,\ldots,x_n/p_n\} \quad (5)$$

このようにして得られる理解結果は、仮説の域を出ておらず、何らかの評価関数によって、その適切

度を判定することが望ましい。式(6)-(8)は、これらの事柄を形式化して表現したものである。

$$h(P,B)=H \quad (6)$$

$$H=\{P[x_1,\ldots,x_n]\theta \mid \text{仮説的修復単語列}\}=\{H_1,\ldots,H_m\} \quad (7)$$

$$e(H)=H' \quad (8)$$

以上の式において、h: 仮説生成関数、H：仮説の集合、e: 適切度評価関数、H':ある評価基準に基づき順序付けされたHである。

7.2 理解結果の適切性の評価

ここでは、複数個得られた理解結果、すなわち、式(7)に示す仮説の集合Hの要素に、ある観点から優先順位を付け、順序集合H'を生成する方法の一つを提案する。この方法は、Hの各要素につき、意味理解に要するコストを計算し、その小さい順に優先順位を付けるというもので、システムにとって、そのようなコストの大小が、理解に関する難易度と比例するとしている。すなわち、「最も理解し易いものが最適の理解結果である」と仮定することになる。

我々は、表層構造の意味解釈結果(理解結果を含む)を、述語論理式として表現している。従って、意味理解に要するコスト（複雑度）は、書き取り結果である表層構造（理解結果を表層構造に変換したもの）から表層依存構造を経て意味構造（述語論理式）に至る各処理過程におけるコストの総和として表現されるべきであろう。このような考えに基づき各過程のコストを分析してみよう。ただし、以下では、書き取り結果を単に出力文章と呼ぶ。

（1）表層構造・表層依存構造間の変換コスト

表層依存構造へ変換するコストは、大まかに出力文章を構成する単語（名詞）の数（$N_O$）に比例すると考える。

（2）表層依存構造・意味構造間の変換コスト

出力文章を構成する単語の意味記述（の結合操作部）に基づく意味構造生成処理では、主に、述語論理式間における項の統合処理（ユニフィケーション）の発生回数（$U_O$）が、省略概念補完処理では、そのような回数（$U_S$）に加えて、補完概念を挿入する回数（$N_S$）がコストとしての考慮対象となる。

以上の考察結果に基づき、意味理解に要する総コスト（$C_T$）を次式(9)で与える。

$$C_T=N_O+U_O+U_S+N_S \quad (9)$$

表層的には、$U_O$、$U_S$および$N_S$は、それぞれ、依存構造における依存関係(係受け対)の個数(D)、挿入単語数（W）、および挿入自立語数（E）にほぼ対応している。すなわち、$C_T$は、近似的に式(10)で与えられる。

$$C_T=N_O+D+W+E \quad (10)$$

更に、依存構造の性質（$D=N_O-1$）から、$C_T$は次式(11)で近似できる。

$$C_T=2N_O+W+E \quad (11)$$

このような評価方法を用いて、不完全な文章である例7の理解結果に関して優先順位付けを試みた結果が表1である。

（例7）父親・$X_1$・自動車・$X_2$・学校・$X_3$・通勤

表1によると、最適なものは、S1に、最も不適なものは、S3およびS4に対応する理解結果となった。

8. むすび

言語理解処理主導型の音声理解システムの構想を述べた。解決すべき今後の課題は以下のようである。

（1）常識の抽出および記述の継続

特に、世界に依存する長期記憶に相当するものが不足している。

（2）選択公準の抽出および記述の継続

今回呈示した局所的および対局的選択公準は、全て排除型のものであるので、積極的に採用する型の選択公準を用意する必要がある。

（3）選択公準の適用条件の定量化

選択公準を一律に適用するのではなく、無意味度や冗長度などの程度により適応を制御するような機構を考える。

（4）状況の記述

特に、言語的脈絡によって話題が何であるかを把握し、現在の発話の内容的評価および次の発話を予測するような機構を備えるのが望ましい。

（5）適切度評価関数

評価関数eのパラメータとしては、現在、補われた概念の個数や、論理的概念表現に出現する項間のユニフィケーションの回数などを採用しているが、単語の音としての特徴（長さなど）も積極的に取り込むべきであろう。

なお、本研究の一部は、文部省科学研究費補助金（重点領域研究 05241105）によっている。

## 参考文献

[1]横田将生他:自然言語理解システムIMAGES-Iの出力合成過程について,信学論,J70-D,11,pp. 2267-2272(1986).

[2]横田将生他: 視覚化された概念モデルに基づく自然語の意味解釈について,信学論,J63-D,5,pp. 417-424(1980).
[3]横田将生他:自動理解処理を目的とする退院サマリの体系的分析,情処学論, 29-12,pp.1217-1225(1988).
[4]横田将生:人間の自然言語理解に関する一つの心理実験,信学論,J71-D,10(1988).
[5]横田将生他:自然言語理解システムIMAGES-II,信学論,J74-D-II,9,pp.1243-1254(1991).

表1　理解結果の適切度評価*

| 文 | θ | $N_O$ | W | E | $C_T$ | P |
|---|---|---|---|---|---|---|
| S1 | {$X_1$/が,$X_2$/で,$X_3$/に} | 7 | 3 | 0 | 17 | 1 |
| S2 | {$X_1$/の所有する,$X_2$/で,$X_3$/に誰かが} | 10 | 6 | 2 | 28 | 3 |
| S3 | {$X_1$/の経営する,$X_2$/に関連する,$X_3$/に誰かが} | 11 | 7 | 3 | 32 | 4 |
| S4 | {$X_1$/の所有する,$X_2$/のある,$X_3$/に誰かが} | 11 | 7 | 3 | 32 | 4 |
| S5 | {$X_1$/が,$X_2$/に関連する,$X_3$/に} | 8 | 4 | 1 | 21 | 2 |

*$N_O$：入力単語数，W：挿入単語数，E：挿入自立語数，$C_T$：概念構造の複雑度（$=2N_O+W+E$），P：適切度の順位。

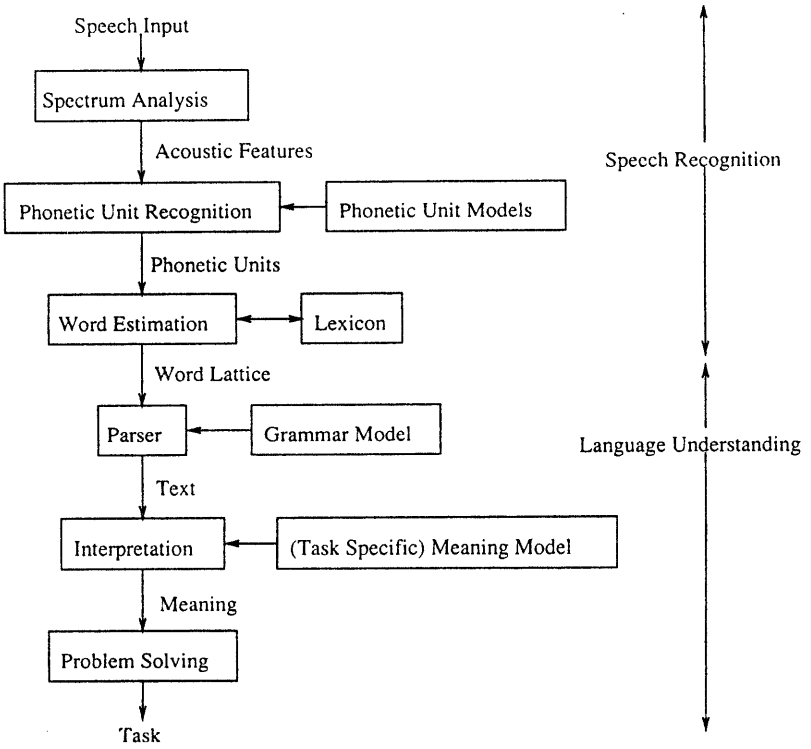


図1　言語処理主導型音声理解システムの概念図

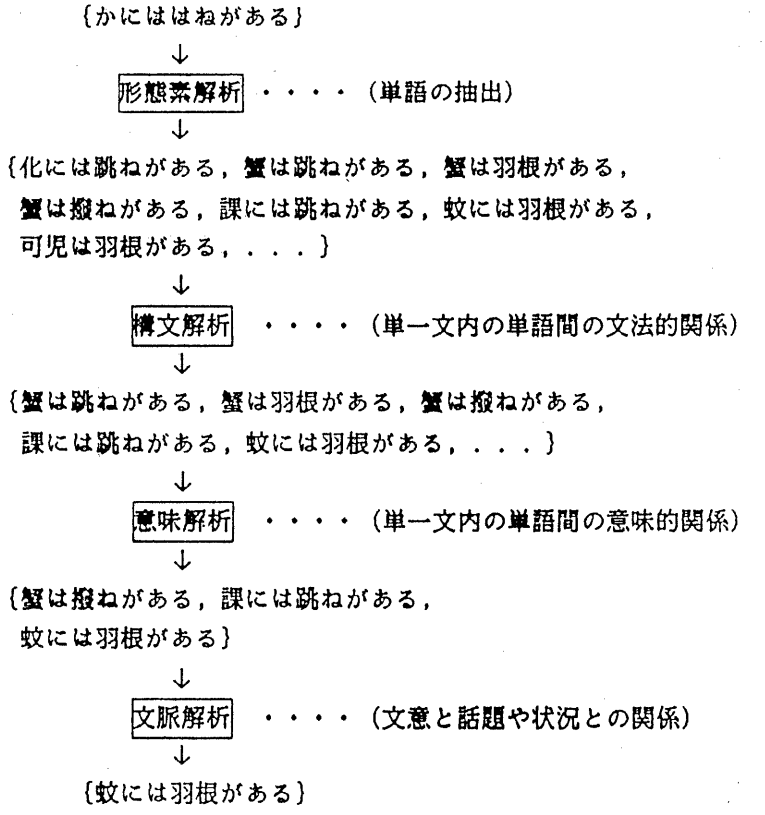


図2　自然言語理解処理の基本型