音声言語情報処理 6－8－1
（1995．5．25）

パネル討論：話し言葉の文法構築は可能か？

中川 聖一

（豊橋技術科学大学・情報工学系）

音声言語によるマンマシンインターフェースを実現するには、書き言葉ではなく話し言葉を機械が理解できなければならない。では、話し言葉の文法を構築することは可能であろうか、また、文法の構築なしで理解可能だろうか議論する。パネラーは、言語学の立場から、金水敏先生、言語工学の立場から伝康晴氏、音声言語工学の立場から竹沢寿幸氏と伊藤克亘氏である。まず、初めに私見を述べさせて戴くことにする。

1．はじめに－－－文法とは－－－

言語モデルの作成法は、人間が規則を作成する方法と機械によって規則を自動学習する方法に分けられ、さらに、書き換え規則を大量データを基に確率的に表現するか否かによって分類される（表1参照）。本論では、議論を明確にするために、人間が明示的に説明できる規則の集合からなるものを文法と呼び、たとえ有限集合の規則で表現できても、その意味が明示的に説明できないものや、抽象化の度合いの小さいものは文法でないとする。例えば、n-gram などのコーパスから機械的に得られる確率モデルは文法でないとする。音声認識用の言語モデル・文法（非文の排除、illformedness の扱い、パープレキシティの減少）と言語解析・理解用の言語モデル・文法（入力文は少々の誤りを除いて正しいと仮定）、文生成用の言語モデル・文法（分かりやすい文のみ生成）は、それぞれ目的が異なるので別々のモデル・文法でよいと考えられる。

2．話し言葉

人間同士のコミニュケーション手段は、音声言語が中心である。身振り、表情も重要な手段だが、電話を介したコミニュケーションを考えれば、音声言語が第一義であることがわかろう。伝達したい概念・意図は曖昧模糊とした場合が多く、この場合でも音声言語として表出しなければならない。幸いに音声言語ほ文字言語と違って韻律情報も含んでおり、文字言語よりもより正確に（曖昧模糊とした）情報が伝達できよう。しかし、我々は自分の言いたいことをなかなか正しく音声言語では表現できないし、同じ概念でもいろいろな表層表現がある。即ち、多対多の写像になっている。これは離散シンボルである文字を発声したり書くと音声パターンや手書きパターンとして表現される1対多の写像よりも複雑である。このことから、いくら、音声言語を正しく理解しようとしても限界があるのは明らかである。まして、一文毎の文を形式的に扱おうとすること自体、無理なことであると思われる。書き言葉ではなるだけあいまいさのないように表現しようとするが、リアルタイム・オンライン生成を基本とする話し言葉ではこのあいまいさを対話によって解決している。

（ｼﾝﾎﾞﾙ）→ 音声・文字ﾊﾟﾀｰﾝ →（ｼﾝﾎﾞﾙ）
1 ： 多 ： 1
（a）音声認識・文字認識

意図→(ｼﾝﾎﾞﾙ)→音声ﾊﾟﾀｰﾝ→(ｼﾝﾎﾞﾙ)→理解
1(多?)：多
1 ： 多 ： 1 ：1(多?)
（b）音声言語理解

図 1 音声認識と音声言語の理解

音声言語によるマンマシンインターフェースを構築しようとする我々工学者は現在の人工知能研究と同様に、計算機で扱える世界・土俵上で、現象を把握しようとする。あるいはこのような世界に写像して、あるいはこの

ような世界のみを対象としている。これは、toy world と呼ばれているもので、徐々に real world へ対象を広げる方向に研究が進んできたと言ってよい（図 2 参照）。

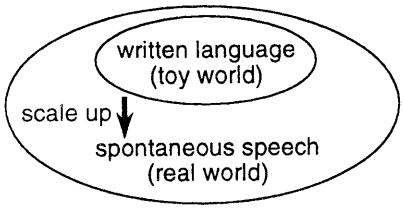

図 2 音声言語処理
(toy worldから real world へ)

3．音声言語と文字言語

音声言語による対話は、人にとって最も自然でかつ多様性、融通性と深さに富むものであり、しかもその対話能力は、人が生まれながらに持っており、自律的に獲得されていく自然でかつ基本的な能力である。一方、文字言語は明らかに後天的に学習・獲得されるものであり、しかも音声言語と比較的一対一に対応していることに特徴があり、このことが、文字言語を音声化したものが音声言語であるという短絡的な先入観による誤解も生じる。音声言語と文字言語は文字言語をサイン言語と同程度に同じでありかつ異なっているととらえる見方も必要であろう。

さて、文字言語には「書きことば（文語）」と「話しことば（口語）」がある。もちろん音声言語は話しことばに対応しているが、言語学者の研究対象はもっぱら書きことばに限られてきたきらいがある。工学分野に限っても同様で、書きことばを対象とした研究（自動抄録、機械翻訳など）が中心であり、最近やっと音声言語による人間と機械とのインターフェース技術が注目されるようになってきた。

機械処理の立場から言えば、音声言語と文字言語の一番大きな違いは、情報源がアナログ波形であるか離散シンボルかである。音声言語特有の現象として、間投詞、助詞落ち、倒置の多用、言い間違い、言い直し、言い淀みの出現、多様な言い回しなどがあり、文字言語と比べて機械にとっては非常に扱い難い代物である。また、声質・韻律という形で個人性、モダリティや感情を伝達できるのも特徴的である。

4．話し言葉の文法は構築できるか

多様な言い回しや表現が存在する話し言葉は、相手の反応を見ながらオンライン的に発話されるもので、これを受理したり生成する文法を構築することは極めて難しい。手書き文字や音声の認識アルゴリズムを明示的に構成するのと同等に困難と考えられる。

例えば、ARPA の ATIS のタスク（語彙数2000～3000単語）の trigram によるパープレキシティ（2のエントロピー乗）は20前後、20000語彙の WSJ タスクで150前後である。文法で表現すれば、数倍は大きくなると思われる。また、我々の作成した小さなタスクの話し言葉用文法（パープレキシティ約70）から、ランダムに文を作成すると、その約8割が非文であった。すなわち、工学的に扱うためには、世界・対象を限定しながら文法を構築し、表現できない部分は例の集合や機械的に構成した文法で対処せざるを得ない（表 1, 図 3参照）。

このような試みを繰り返しながら、よりカバーレイジの大きい文法を構築していかざるを得ない。その都度、工学的に応用できる分野を見つけて未完成な技術を実用化していく努力が必要であろう。

表 1 言語モデルの比較
構文解析・書き言葉用：解析木の良否
音声認識・話し言葉用　：ｴﾝﾄﾛﾋﾟｰの大小

ｱﾌﾟﾛｰﾁ	対象	人間による規則生成	機械による規則(ﾓﾃﾞﾙ)生成
非統計的	構文解析	△	×
	音声認識	△	×
統計的	構文解析	◎?	○
	音声認識	◎	◎?

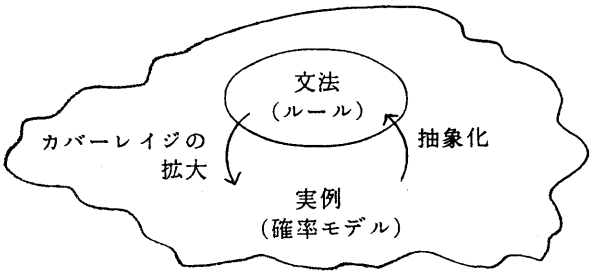

図 3 文法と実例

話し言葉の文法構築は可能か?

伝 康晴
ATR音声翻訳通信研究所

1 はじめに

近年，音声処理と言語処理の境界領域として，話し言葉の研究が盛んである．しかし，話し言葉では，言い淀み，言い直し，言い誤り，構成素の欠落などが多く見られ，その機械的処理を困難にしている．これは，構文解析のための文法や音声認識で用いる言語モデルが話し言葉ではうまく構築できないことが主な原因である．本稿では，話し言葉の文法はいかにして構築できるかについて考える．

2 2つの文法

そもそも，文法とは何であろうか．理論言語学では，文法を「人間が持つ言語能力」と位置付けている．これを文法Aと呼ぼう．すなわち，文法Aとは，人間が言葉を理解/産出する際に用いている知識の体系のことである(よって，話し言葉の文法と書き言葉の文法の区別はない)．一方，自然言語処理では，文法を「対象言語の文に見られる規則性」と位置付けている．これを文法Bと呼ぼう．すなわち，文法Bとは，ある言語に属する文の性質を記述する体系のことである(よって，話し言葉の文法と書き言葉の文法の区別がありうる)．

もし，(1) 人間の言語処理が言語能力としての文法A以外の要因の影響を受けておらず，かつ，(2) 文法Bが人間の言語をモデル化するのに十分な道具立てを備えたモデルに基づいているなら，文法Aが生成する言語Aと文法Bが生成する言語Bとはともに現実の言語と一致する．しかし，実際には(1), (2)は成り立たない．人間の言語，特に話し言葉においては，記憶容量の制限や発声器官のエラーなどによって，文法Aが産出した文が必ずしもそのままの形では表出されないし，また，自然言語処理が文法Bのモデルとして用いている遷移ネットワークや文脈自由文法(あるいは，これらを確率的にしたもの)は現実の言語をモデル化するのに十分とはいえない．その結果，言語A，言語B，現実の言語はすべて異なったものとなるのである(図1)．

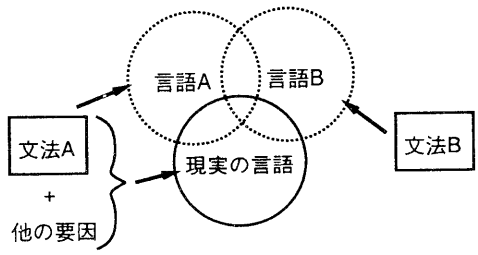

図1 2つの文法と言語の関係

3 話し言葉を扱う2つのアプローチ

自然言語処理で話し言葉を扱うには，大きく分けて2つのアプローチがあると思われる．1つは，図1の左側に記された人間の言語処理を模して，文法だけで現実の言語をすべてカバーするのではなく，他の機構を持ち込むことによって，文法がカバーする言語と現実の言語とのずれを補おうというものである．Fitted parse [6]や緩和法 [7]などはこういった試みと考えられる．もう1つは，図1の右側で，文法モデルを拡張することによって，現実の言語を正しくカバーする文法を作ろうというものである．優先意味論 [4]やアブダクションに基づく解釈 [5]などはこういった試みと考えられる．

一見すると，前者のほうが正直な方法のように見えるが，必ずしもそうとはいえない．なぜなら，人間の言語処理における「他の要因」は極めて複雑で，モデル化が困難であり，それゆえ，前者のアプローチでいう「他の機構」は実際にはこれをモデル化したものにはなっていないからである．このような機構は，通常，ヒューリスティクスに基づいて設計されるが，これは，生成文法や隠れマルコフモデルといった形式的理論に比べるとかなりいぶかしいものである．例えば，「ほん，翻訳を入れます」という，通常の文法では解析に失敗しそうな文(非

文)が与えられたとき，文法をどのように緩和するかにはいろいろ選択肢がある(最初の語「ほん」を単に捨ててしまう，あるいは，助詞を補って「本に」という助詞句にして動詞にかけるなど)が，どれを選ぶかはアドホックな優先規則によってしか決まらない．この欠点は，現実の言語においては，文法の緩和の仕方に関してもある種の規則性や偏りが見られるはずであるにも関わらず，それが適切にモデル化されていないことに起因する．

一方，後者のアプローチは，適格文と非文とを区別していないので，適格文の文法(文法A)の存在を否定する反言語学的なアプローチのように見える．しかし，そうではない考え方もある．例えば，アブダクションに基づく解釈[5]では，ある統語的制約が実際には成立していないときにも，それが成立すると仮定してしまうことによって，非文を適格文と同じように扱うことができるが，これは，このモデルにおいて適格文と非文との区別が全くないということではない．つまり，統語的な仮定を作らずに解析できる文が適格文であり，作らないと解析できない文が非文である．よって，このモデルにおいては，文法Bが文法Aを包含していると考えられる(ただし，文法Aが文法Bの部分集合になっているわけではなく，適格文と非文との区別はあくまでも計算の内容(仮説を作るか否か)によって決まることに注意せよ)．このモデルにおいては，「文法A＋他の要因」全体を，(直接的にではないにせよ)ある種の形式的体系によってモデル化しようとしており，文法Aに属さない要因の規則性や偏りをうまくモデル化できる可能性を持っている．

4 日本語の話し言葉の文法

ここでは，日本語の話し言葉の文法について，前節の2番めのアプローチにそって，筆者自身が行なっている研究について簡単に紹介する[1][2][3]．

このモデルでは，文法として，統語的な規則だけでなく，意味的な規則や言い直しの性質などに関する規則を記述している．このモデルに「ほん，翻訳を入れます」という文が与えられると，3つの文節「ほん」「翻訳を」「入れます」相互の間にどのような係り受け関係が可能かということが文法を用いて計算される．例えば，「ほん」と「翻訳を」の間には発声上の問題によって生じた言い直し(「ほん」まで言いかけて「ほんやくを」と言い直した)という関係が考えられる．あるいは，「ほん」と「入れます」の間に，助詞「に」を欠いているが，場所を表す関係があるとも考えられる．これらは，いずれも確証のない仮説であるが，もっともらしさを表す尤度がついている．仮説の尤度は，別に用意した訓練例をあらかじめ解析しておき，その仮説と似たものがどのくらい頻繁に現れるかを調べることによって与える(詳細は[3]参照)．おのおのの仮説の尤度の大小を比較することによって，もっともありえそうな解釈を選択することができる．

このモデルでは，係り受けの意味的な整合性，助詞の落ちやすさ，さまざまなパターンの言い直しの起こりやすさなどが統計的にモデル化されており，「文法A＋他の要因」全体をモデル化しようという試みといえる．

5 おわりに

話し言葉の文法をどうとらえるかについて述べ，それがいかにして構築できるかの一例を示した．筆者のモデルは，話し言葉を扱う第2のアプローチに属しており，今後，語用論的な要因，文脈的な要因まで取り込める可能性を持っているが，音声処理との統合を考えた場合，統語規則以外のものを駆使するモデルと必ずしも相性がよいとはいえず，fitted parseのような第1のアプローチとの併用も検討すべきであろう．

参考文献

[1] 伝, 飯田: 情報伝達の観点から見た日常会話文の解析手法, 「自然言語処理の新しい応用」シンポジウム論文集, 電子情報通信学会・日本ソフトウェア科学会, pp. 40–49 (1992).

[2] 伝: 音声会話文法の定式化について, 人工知能学会研究会資料, SIG-SLUD-9302, pp. 33-40 (1993).

[3] 伝: 制約と統計に基づく自然な発話の解析, 言語処理学会第1回年次大会発表論文集, pp. 41–44 (1995).

[4] Fass, D., and Wilks, Y.: Preference Semantics, Ill-Formedness, and Metaphor, *Computational Linguistics*, Vol. 9, No. 3–4, pp. 178–187 (1983).

[5] Hobbs, J.R., Stickel, M.E., Appelt, D.E., and Martin, P.: Interpretation as Abduction, *Artificial Intelligence*, Vol. 63, No. 1–2, pp. 69–142 (1993).

[6] Jensen, K., Heidorn, G.E., Miller, L.A., and Ravin, Y.: Parse Fitting and Prose Fixing: Getting a Hold on Ill-Formedness, *Computational Linguistics*, Vol. 9, No. 3–4, pp. 147–160 (1983).

[7] Weichedel, R.M. and Sondheimer, N.K.: Metarules as a Basis for Processing Ill-Formed Input, *Computational Linguistics*, Vol. 9, No. 3–4, pp. 161–177 (1983).

話し言葉の文法構築を目指して

竹沢 寿幸
ATR 音声翻訳通信研究所

1 まえがき

音声翻訳ないし音声対話システムの構築を目指して、自然で自発的な発話を対象とする連続音声認識の研究を進めている。その一環として、音声認識用日本語文法の検討を行なっている。音声認識における言語モデルの１つとしての日本語文法には、次の利点がある。

- 単語や品詞の統計的な連接情報(典型的にはバイグラムやトライグラム)のみでは、発話としてあり得ないような系列が生成されることがある。統語的な制約を利用すれば、統語的にあり得ない系列を除去することができる。
- 音声翻訳ないし音声対話システムを構成しようとする場合、音声認識系からの出力としては、音素系列よりも単語系列、単語系列よりも部分木系列が望ましい。音声認識用日本語文法を利用すれば、部分木仮説を生成することができる。

2 N-gram モデルと構文規則ベース型

N-gram モデルは統計的な言語モデルの典型である。コーパスさえあれば、自動学習(automatic learning)が可能である。しかし、訓練(training)のために大量のテキストが必要である。新聞記事のような書き言葉ではなく、話し言葉を対象とすると、大量に書き起こしテキストを集めるには莫大な手間およびコストがかかる。しかも、領域(ドメイン)や課題(タスク)に依存してしまう。

一方、構文規則ベース型における構文規則は、領域や課題にあまり強く依存せず開発可能と期待できる。語彙項目(lexical entry)さえ変更すれば、複数の課題／領域にポーティング可能である。ヒューリスティックあるいは統計に基づくスコアを付与することは容易である。

したがって、音声翻訳ないし音声対話システムのフロントエンドとしては、構文規則に基づき部分木系列をスコア付きの仮説として出力する、音声パーザ(speech parser)が有用であろう。

ただし、語彙サイズが大規模となったり、発話様式が激しく変化する状況に対応しようとすると、音声認識過程で直接的に構文規則を言語モデルとして利用する手法が最適かどうかはわからない。探索手法と言語モデルの組み合わせ方に関しては、対象とする応用分野(語彙サイズや発話様式を規定する条件)に応じて、さまざまな手法を精力的に研究する必要がある。

3 自然で自発的な発話の言語現象

音声翻訳ないし音声対話システムを構築するという立場から言語現象の分類調査を行ない、文献[1]に結果を報告した。頻度の多いものを表1に示す。

表 1: 不適格な言語現象と頻度(上位のもの)

265：言い淀み(文頭)	32：中止文(接続助詞終止)
131：気づき(文頭)	27：提題助詞の欠落
84：助動詞の欠落	25：連体助詞の欠落
71：言い淀み(文中)	23：だ型文
53：格助詞の欠落(を)	22：つづり
47：特殊構造(箇条発話)	20：直示
46：応答(文頭)	19：サ変名詞述語
37：述語の欠落	18：後置詞句の名詞修飾
36：連続文	18：融合文

このような不適格な現象を文法規則でどのように扱えばよいか、**特殊構造(箇条発話)**の例を挙げて説明する。**特殊構造(箇条発話)**とは、ホテルや列車の予約の際に、確認のために予約項目を読み上げるものである。

例文１ 鈴木和子様、八月の十日から十二日まで、シングルルームシャワー付き二泊ですね。

例文２ はい、ワシントンディーシーへ、メトロライナー十二時二十分発、ユニオン駅でよろしいですね。

表 2: 音素複雑度 (perplexity) の比較

文法	テストセット	音素複雑度
適格な文法	文節区切りの文	2.49
適格な文法	文	3.23
言い淀み (間投詞) つき文法	文節の位置に言い淀み出現を許容しない文	3.39
言い淀み (間投詞) つき文法	文節の位置に言い淀み出現を許容する文	4.06

これらの発話は、外部の状況がそのまま言語に反映されているといったもので、あらかじめどのような順序で出現するかを予測するのは難しい。しかも、統語的なカテゴリのみで規則化することは困難である。さらに、ポーズで区切って発話される傾向がある。そこで、無理な規則化を避け、部分木として出力する。なお、この構文規則は一般的に作成しているので、ある特定の課題／領域に特化するのは容易である。つまり、音声対話システム研究のベースとなるものと考えている。

4 ポーズの調査

ポーズはもともと息継ぎのために置かれるもので、発話の過程にとってはあくまでも外的なものである。しかし、いったんその過程に入り込むと、統語的・意味的構造の制約を受けないわけにはいかない。

一方、聞き手は、ポーズで区切られたまとまりを手がかりにして発話を理解している。ポーズをすべて取り除くと、聞き手はまったく発話を理解することができなくなることを立証した実験もある [2]。

そこで、文法開発者が構文木の形に注意してポーズ位置の観察を行なった。文法開発者の主観によりポーズの前後に何らかの構文木が構成できると判定できる箇所に存在するポーズをマークアップした [3]。その結果、英日逐次通訳者の発話を除けば、自然で自発的な発話に対して、ポーズを考慮して、部分木を単位とする日本語文法が設計開発できそうな感触が得られた。

5 予備実験

自由で自発的な発話に対して、文全体をカバーする文法を書くことは難しい。一方、音声言語解析 (格解析) 部への入力として部分木を採用することは魅力的である。そこで、部分木を出力する音声認識系が有効ではないかと考え、朗読音声を対象に予備実験を行ない、文献 [3] に結果を報告した。その結果は、部分木を出力する音声認識系が有効そうであることを示唆している。

しかしながら、もし文法を音声認識過程で直接的に利用しようとすると、慎重に文法およびメカニズムを設計しなければ、急激に音素複雑度が悪化する。例えば、既存の「適格な文法」に言い淀み (間投詞) を組み入れた場合に、どの程度、音素複雑度が変わるか比較した結果を表 2に示す。言い淀み (間投詞) としては、9 種類の典型的なものを扱った。

6 今後の展望

設計および試作の済んだ音声認識用日本語文法 [3] を用いて、連続音声認識の評価実験を進める。1 万語程度までの大語彙を意識した検討を行ないたい。音声認識用日本語文法のみにこだわらず、確率・統計・クラスタに基づく言語モデルとの組合せも試みる。一方で、音声認識過程での探索手法についても精力的に研究を進める必要がある。さらに、音声対話システムを考慮した音声認識手法の研究を行なう予定である。音声対話システムの意味表現や、音声認識結果を利用して、構文規則を (半) 自動的に保守するような研究にも興味がある。

7 むすび

話し言葉の文法構築を目指している立場から、意見を述べた。音声翻訳ないし音声対話システムのフロントエンドとしては、構文規則に基づき部分木系列をスコア付きの仮説として出力する、音声パーザ (speech parser) が有用であろう。

参考文献

[1] 竹沢寿幸, 田代敏久, 森元逞: "音声言語データベースを用いた自然発話の言語現象の調査", 人工知能学会 言語・音声理解と対話処理研究会 (第 10 回), SIG-SLUD-9403-3, pp. 13-20 (1995-02).

[2] 杉藤美代子: "談話におけるポーズとイントネーション", 日本語と日本語教育 第 2 巻 — 日本語の音声・音韻, 明治書院 (1993).

[3] 竹沢寿幸, 田代敏久, 森元逞: "自然発話の言語現象と音声認識用日本語文法", 情報処理学会 音声言語情報処理研究会, SLP-6-5 (1995-05).

音声認識に役立つ？ 文法

伊藤克亘
電子技術総合研究所
email: kito@etl.go.jp

1 はじめに

本稿では、音声対話システムの構築を通して感じた音声認識における文法のあり方について、とくに実世界と文法のかかわりといった視点から論ずる。

2 音声認識と文法

現在の音声認識システムでは、文法や辞書などの(言語的な)知識(以下では、まとめて文法とよぶ)を利用して、発話される可能性のある仮説を作り、その仮説にしたがって音響的な標準パターン(以下では、音韻モデルとよぶ)を連結して、発話に近い仮説をさがす。この関係を模式的にあらわすと以下のようになる。

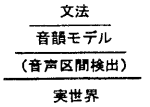

図 1. 現行の音声認識システムにおける文法

実際には、まず、入力のうち音声の存在する区間だけを検出して、その区間だけを対象に認識をおこなうことが多いので、上の図では検出の部分も加えてある。

ごく平凡なシステムでは、音韻モデルとして音素モデル、文法として、音素を終端記号とした音素列を生成するような辞書/文法を用いることが多い。つまり、辞書としては、語を音素系列として表わしたもの、いわゆる文法としては、どの語とどの語が連結するかを表現した程度のものを使っており、自然言語処理の分野でいうと形態素解析程度のことしかしていないのが現状である。

その程度のものしか使えない理由は、最初にも述べた通り、可能な限りの仮説を生成する必要があるからである。多くの意味処理のように、例外的な不適格な表現(名詞と動詞の組み合わせなど)を除くような処理は、いったん認識して複数の候補を求めておいて、後処理的に利用することが多い。本稿では、後処理的に利用する文法は考察せず、直接音声認識に利用する文法についてのみ考える。

3 現行の音声認識システムの限界

現在の音声認識システムは前節で述べたような構成なので、検出部分で失敗したら、捨てられた部分に対して何も認識しない。また、ほとんどのシステムでは、切り出された部分は、文やそれに準ずる表現であると仮定しているため、いい直しやいい誤りといったものに対応することは難しい。さらに、音韻モデルだけで発話をとらえているために、音韻以外の要素は捨てられている。

こういった枠組では、静かな部屋で、アナウンサーなどが朗読した音声であればかなり認識できるのかもしれないが、音声対話システムがめざすような話し慣れていない人を相手にする場合には、たいてい発話の半分も扱えない。

文法との関わりでは、上記の限界に対して、「話し言葉」の表現に関する知見/文法規則が十分でないから駄目である、という問題点を第一にあげる人が多い。では、たとえば、「えー」とか「あのー」といったつなぎ語(filler)を含んだ発話を認識するためには、文頭や文中に「えー」や「あのー」が挿入されるような規則を用意すれば扱えるのだろうか? 現在の枠組では、仮にそういう規則を用意しても、余り精度良く扱えない。なぜ、うまくいかないかを考えるためにも、次節では、現在の音声認識の枠組で捨象されている事柄について考えてみる。

4 実世界と言葉をつなぐ様々な要素

現行の音声認識システムがすてている要素を列挙してみる。

- 音声の時間
- 音声の高低・強弱
- 音以外の情報
- 雑音

現在の音声認識システムでは、音声区間を検出し、それ以外の無音とみなした区間を捨てている。さらに、その音声区間は全体でひとつの文であるとみなして認識している。したがって、時間的な概念としては、せいぜい、音韻の順番や発話の順番くらいしか考慮されているとはいえない。しかし、人間がある発話

を聞いたとき句点を意識するようなところに、物理的にはほとんど休止がなかったり、言い間違いや言い直し・つなぎ語の前後には、かなりの割合で休止が生じるなど、無音区間も言語的に重要な情報を担っていると考えられる。

つなぎ語と同じ役割を果たす現象として、「井の頭通りのーところで」のように、語のある音韻を伸ばす(ここでは、「のー」のところ)発話がある。現在の枠組では、「のー」の部分は通常の発声と違うため音韻認識の精度が下がってしまうだけで、この現象から得られる有効な情報を生かす方法はよくわかっていない。そもそも、「えーと」などのつなぎ語と同様、ここで便宜上「のー」表記した部分も、その性質上音素とはいい難い面もあるため、音素表記したり、音素として認識することにどのくらい意味があるのかわからない。

また、時間的な概念としては、テンポ・リズムなども、発話のやりとりを円滑にする上で重要であると考えられる。しかし、現状では、これらをどう文法に反映させるか、まるでわからないため、システムの応答の時機はぎこちない不自然なものになってしまっている。

音の高低・強弱も、語句の強調など様々な情報を担っていると考えられる。このうち、かかり受け構造と関連するような情報については、これまでの文法との親和性が高いため、かなりの試みがあるが、ある語句が強調されていることを認識するためにはどうすればいいのか、とか、それをどのように文法上で反映させるか、などについては、ほとんど手つかずの状態である。

音韻以外の情報としては、二種類に分類できるだろう。ひとつは、環境から得られる情報で、たとえば、自分の目の前に話し相手があらわれたか、とか、話している相手は誰かなどに関する情報である。ユーザがシステムの前に現われたら、システムからユーザに話しかけるような機能を持たせる場合には不可欠な情報である。

もうひとつは、対話相手から与えられる情報で、たとえば、うなづき・身振り・手振り・表情などである。ある質問に「はい」と答える場合でも、うなづきながら答えるのと、とまどった表情を見せながら答えるのでは、おのずから解釈は違ってくる。

唇の形を認識して音韻の認識を補助しようとする試みがある。唇の形は視覚的な情報であるが、音韻的な情報でもあるため従来の枠組との親和性は高いと考えられる。

音声認識では、ほとんどの場合、雑音は排除されるべきものである。しかし、言語理解といった観点では、雑音も重要な情報源であるという気もする。たとえば、電話がかかってきたときに、背景に電車のアナウンスが聞こえていれば、当然、そこでの発話は、駅にいることを前提に認識/理解されることになる。また、発話者が発する、笑い声のような雑音も、発話や対話をとらえる上では、捨ててはならない情報ではないだろうか。

5 まとめ — 理想の音声認識システム

前節で述べたような情報を積極的に利用するためには、音声認識システムをどのように構築すればよいだろうか?たぶん、以下に示したようなシステム構成をとる必要があるだろう。

図 2. 実世界で使える音声認識システムにおける文法

つまり、音韻以外のさまざまな情報によって、実世界との接点を持つ必要があるのではないだろうか。このようなことを書くと、さまざまな文法の枠組で、ある種の手続きを埋め込むような試みがされているので、そういった枠組を利用すれば簡単にできるなどという声が聞こえてくる気がする。ここで問題にしたいのは、そういった抽象的な話ではなく、たとえば、時間を考慮するという場合には、250ms の無音区間があれば、句点とするといった類いの具体的な話である。実世界での使用に耐えるシステムの文法には、そういう具体的な終端記号が多彩に用意されなければならないだろう。また、本稿であげた、捨ててしまっている情報の多くは、従来の言語処理から見れば超文節的 (suprasegmental) な情報が多いため、従来の文法の記述法や解析方法とは相容れない部分もあるだろう。

逆に、現在のままの音韻の精度では、普通の自然言語処理の手法をそのまま持ってきて、音韻だけで実世界との接点を持つような枠組は、これまでも多くの手法がさしたる成果をあげなかったように、実りある結果をえることはできないのではないだろうか。

音声言語情報処理 6－8－5
（1995. 5. 25）

談話標識の諸レベル

—談話管理理論の視点から—

金水 敏
神戸大学 文学部
e-mail: kinsui@icluna.kobe-u.ac.jp

1 談話管理理論

話ことばの文法をどのように構築するかという問題については、さまざまなアプローチが可能である。ここでは、田窪 (1992)Takubo & Kinsui (1992) 等で主張されてきた「談話管理理論」をひとつのヒントとして提示したい。

談話管理理論では、発話を心的データベースへの操作指令と考える。そして発話を大きくデータ部とデータ管理部に分割する。データ部は主に名詞、動詞、形容詞、格助詞、状態副詞、ある種の助動詞等からなり、概念的要素や命題の設定を行う。データ管理部はその他の助詞、副詞、助動詞、感動詞・間投詞等からなり、データベースに対する複写・検索・推論等の操作指令、ないし操作のモニターとして機能する。

ここでいう心的データベースとは一種の知識・記憶モデルである。長期記憶、短期記憶、そしてこれらを結びつける談話領域によって構成される。発話が直接操作する対象は談話領域である。日本語を中心とした考察の結果、談話領域をおおむね３つに区分することが有効であると考える。それはD領域、I領域、そしてC領域(仮称)である。C領域は発話に先だって種々の計算を行うための領域で、さらに言語的操作を行う場合と非言語的操作を行う場合との違いに間投詞「あの一/えーと」の分布が対応することがSadanobu & Takubo (1993) で述べられている。

D領域とI領域は次のように定義できる。

D領域 長期記憶に結合されている。
直接経験や過去の経験から得られた直接的な情報が収められる。

I領域 作業領域に結合されている。
伝聞、推論、定義的情報から得られた間接的な情報が収められる。

日本語の場合、裸の固有名詞、代名詞「彼/彼女/彼ら」、指示詞「こ/あ」はD領域の対象を指し示し、「[固有名詞] という/って [普通名詞]」等のメタ指示形式や指示詞「そ」はI領域の対象を指し示すと考えられる(T & K, 1992)。

2 談話のモデルと終助詞

Kinsui & Takubo (1993) では、従来の「談話の目的は話し手・聞き手の非共有知識を最小化し、共有知識を最大化することである」というモデルを批判し、「談話の目的は未立証知識(I領域内の知識)を最小化し、立証済み知識(D領域内の知識)を最大化することである」というモデルを主張した。その中で、ある未立証知識を相手の知識によって立証しようとする際の談話上の手段として、次の二つがあることを述べた。

1. 相手に直接聞く。
2. 当該知識を現在立証中であることを相手に見せる。

例えば英語の付加疑問文は1に対応し、日本語の終助詞「ね」は2に対応するとした。また、金水 (1995) では、終助詞「よ」は当該命題を未立証知識として提示する機能を持つと述べ、さらに、「早くしろよ」「早くしてくださいね」等の命令・依頼文に付く「よ」「ね」は、命令・依頼という発話行為を直接遂行するの

ではなく、「私は〜と命令・依頼している」という命題の立証問題として相手に提示する機能を果たすとした。本発表ではさらに「ね」の間投助詞用法についても同様の見方をあてはめることを提案する。

3 間投助詞と"半疑問"

次のような間投助詞用法の「ね」は、文素材(主として"文節")を単位とする「確認」を行っている(益岡, 1992)。

(1) きのうね、ぼくがね、大学に行ったらね、…

この「確認」は、「これこれの文法機能を持ったかくかくという語彙を文の素材としてあなたは適切に聞き取りつつある」という推定に関するもので、すなわち談話コミュニケーションの過程の適切さについての確認であると見ることができる(Clark, 1993)。このように見ると、「ね」は「文素材の受け渡し」「命題内容」「(命令・依頼等の) 発話行為」という異なるレベルにおいて機能していると考えることができるのである。

さらに、近年"半疑問(half question)"と一部で呼ばれている特殊な文音調が観察されている。

(2) うちの近くに図書館の出張所↑ ⟨*pause*⟩ができてね、……

これは、用語すなわち文素材の適切さを相手に直接問いながら談話を進めていく方略と位置づけられよう。

4 結論と展望

ここで述べたように、談話管理理論は、談話標識の諸機能を知識モデルの枠内で理論的・統一的に説明することがすることが可能であり、話しことばの文法構築に一定の展望を与えることができるものと思われる。

ただし最後に述べた間投助詞や半疑問については、出現位置の制約、音調的特徴、文末表現との呼応[1]等、明らかにすべき課題が多い。

参考文献

金水 敏(1995) 「知識モデルに基づく感動詞・終助詞・応答詞類の包括的研究」『音声・言語・概念の統合的処理による対話の理解と生成に関する研究』文部省科学研究費補助金重点領域研究研究成果報告書(研究代表者:堂下修司)

Kinsui, Satoshi and Takubo, Yukinori(1993) "Discourse Management Analysis of some of the Japanese Sentence-final Particles," *INTERNATIONAL SYMPOSIUM ON SPOKEN DIALOGUE -New Directions in Human and Man-Machine Communication- (Proceedings ISSD-93)*

Sadanobu, Toshiyuki and Takubo, Yukinori(1993) "The Discourse Management Function of Fillers- a case of "eeto" and "ano(o),"" *Proceedings ISSD-93*

田窪行則(1992) 「談話管理の標識について」『文化言語学—その提言と建設』(三省堂)

Takubo, Yukinori and Kinsui, Satoshi(1992) "Discourse Management in terms of Mental Domains," 『高度な日本語記述文法書作成のための基礎的研究』平成3年度科学研究費補助金総合研究(A)研究成果報告書(研究代表者 益岡隆志)

益岡隆志(1991) 『モダリティの文法』くろしお出版

Clark, Herbert H.(1993) "Managing Troubles in Speaking," *Proceedings ISSD-93*

[1]尾上圭介氏の個人的談話による。