

## 支援対話に現れる述語接続助詞「けど」の意味解析

熊本 忠彦 伊藤 昭

郵政省 通信総合研究所 関西支所

我々は、ユーザの計算機利用を支援する音声対話システムを構築している。このような音声対話型のシステムがユーザフレンドリーであるためには、ユーザの自由かつ自然な発話を理解できなければならない。そのためには複文の形でなされるユーザ発話をも正しく理解する必要がある。そこで本稿では、話し言葉において多用される述語接続助詞「けど(けれど/けども/けれども)」に焦点を当て、この助詞を対象とする意味解析手法を提案する。

### On Semantic Interpretation of the Postpositional Word "KEDO" which Works as a Conjunctive Between Predicates

Tadahiko KUMAMOTO Akira ITO

Kansai Advanced Research Center, Communications Research Laboratory, MPT

We are developing a dialogue system which helps users to work with computers. For such a dialogue system to be user-friendly, understanding user's spontaneous speech is essential. This implies that the system should understand user speech which is made in the form of a complex sentence. This article focuses on the postpositional word *KEDO* (*KEREDO*, *KEDOMO* or *KEREDOMO*) that works as a conjunctive between predicates, and proposes a semantic analysis method for it.

#### 1. まえがき

我々は、ユーザ (computer user) の計算機利用を支援する音声対話システム (以下、支援システムとよぶ) を構築している [1]。このような音声対話型のシステムがユーザフレンドリー (user-friendly) であるためには、ユーザの自由かつ自然な発話 (spontaneous speech) を理解できなければならない。そのためには複文の形でなされるユーザ発話をも正しく理解する必要がある。

本稿では、話し言葉 (spoken language) において多用される述語接続助詞「けど」に焦点を当て、この助詞を対象とする意味解析手法を提案する。なお、「けど」は、「けれど」や「けども」とともに、述語接続助詞「けれども」の縮約形であるが、本稿ではこれらを区別せずに「けど」と表記する。

広辞苑 [2] によれば、「けど」には「前に述べた事柄と後に述べる事柄とが矛盾する関係にあることを表す。」という役割があると述べられている。また、文法書 [3] でも逆接を表す副詞節を形成する助詞として扱われている。しかしながら、

実際の対話で使われている「けど」には、より多様な意味解釈が成り立っているように思われる。実際、ある国語辞典 [4] には「二つの事柄が並べられる関係を表す。前件にかかわらず、対比的な後件、前件に矛盾する、あるいは予想されぬ後件が起こる関係を表す (逆接確定条件)。ある事柄について説明的な条件を述べる関係を表す。」と書かれているし、また別の国語辞典 [5] には「話の糸口として述べられた事柄をごく軽い気持ちで、次に述べる事柄と結び付けることを表す。なんらかの意味で対比される事柄を続けて言うことを表す。」とも書かれている。これらの説明は「けど」に前件と後件の何らかの関係を表す役割があることを示しているが、その内容は抽象的かつ直観的であり、具体的な処理手順を必要とする計算機処理には応用できない。そこで本稿では、実際の対話において使われた、「けど」を含む複文発話を分析することによって、話し言葉における「けど」の役割を明らかにし、「けど」を対象とする意味解析手法を設計する。

なお、支援システムがサポートするタスク領域

として UNIX 計算機上の X ウィンドウ環境下で動作する電子メール処理プログラム XMH[6] が設定されている。XMH は、電子メールを受け取る/表示する/作成する/送る/保管するといった機能を有し、主にマウスで操作される。

以下、本稿の構成について述べる。2. で「けど」を含む複文発話の収集について述べる。3. で発話文の意味表現形式として発話意図記述を定義し、4. で話し言葉における「けど」の役割を明らかにする。具体的には、「けど」の前件と後件をそれぞれ発話意図記述で表現し、発話意図記述の組合せパターンと「けど」の役割との関係を分析する。また、この分析の結果を利用して、「けど」を対象とする意味解析手法を設計する。5. でこの意味解析手法をベースにした、複文発話の発話意図解析手法について述べる。最後に 6. で本稿のまとめと今後の課題について述べる。

## 2. 「けど」を含む複文発話の収集

我々は、支援システムの代わりに人間コンサルタントが支援するという「対話による支援」実験を行い、ユーザの自由かつ自然な発話 (spontaneous speech) を収集した。そして、収集されたユーザ発話を書き起こすことによって、支援対話データベース (ユーザ 42 人による発話文 855 文を含むテキストデータベース) を作成した [7]。

この支援対話データベースから「けど」を含むユーザ発話文 (全 54 文) を抽出し、ユーザ発話文コーパスを作成した。ユーザ発話文には、日常会話文と同様、間投表現や言い直し、言い淀み、言い換え表現などが現れていたが、これらはすべてコーパス作成にあたって取り除かれた。このような表現の取り扱いには、今後の課題である。なお、助詞落ちや助詞の誤運用、倒置なども観測されたが、これらはそのまま残されている。

また、「けど」の前件、「けど」本体、「けど」の後件からなる三つ組を「けど」事例とよぶことにすると、ユーザ発話文コーパスには、55 個の「けど」事例が含まれていた。これは、複文 54 文のうち 1 文に「けど」が 2 回現れていたためである。

## 3. 発話意図記述の定義

発話文の発話意図は命題とモダリティによって表現される [8]。命題 (タスク領域上で定義可能な「もの」や「行為」、「状態」) を表現する記述は、「発話命題」と呼ばれ、フレーム表現で記述される。また、モダリティ (発話命題に対するユーザの

表 1 発話意図タイプと支援要請の種類

発話意図タイプ	支援要請の種類
手法	手法に関する質問
概念	概念に関する質問
属性値	属性値に関する質問
属性値:OK	事象そのもの 未来事象の正当性
真偽値	真偽値に関する質問
真偽値:OK	事象そのもの 未来事象の正当性
真偽値:NO	未来事象の禁止性
真偽値:EQ	その他
信念	事実等の表明
ゴール	タスクゴールの表明
対話開始	対話開始の合図
対話終了	対話終了の合図

心的態度) を表現する記述は、「発話意図タイプ」と呼ばれ、表 1 に示した 12 個のシンボルの中から 1 つ選ばれる。従って、1 つの発話意図は、

(発話意図タイプ 発話命題)

という形で記述される [9]。我々はこの記述を発話意図記述と呼んでいる。なお、表 1 において、未来事象の正当性に関する質問とは「～したいんですか」のようなある事象の容認可能性を問うことによって許可を求める文のことであり、未来事象の禁止性に関する質問とは「～してはいけませんか」のような禁止の形式で問うことによって許可を求める文のことであり、また、事象そのものに関する質問とは XMH 上で生じた事象 (出来事/状態) に関する質問のことであり、その他のことに関する質問とは XMH 上のある状態の正当性に関する質問のことであり、

## 4. 「けど」事例の分析に基づく意味解析手法の設計

「けど」はその前件と後件の何らかの関係を表している。そこで、前件が有する発話意図と後件が有する発話意図の関係を調べるために、前件および後件から生成される発話意図記述の関係を調べた。具体的には、前件の発話意図タイプと後件の発話意図タイプとの組合せパターンにより、「けど」事例 (55 個) を分類した。この結果を表 2 に示す。このとき、後件から得られた発話意図記述に対しては、その発話命題の有無も調べた。これは、文献 [10] で示したように、発話命題が生成さ

表2 発話意図タイプの種類と発話命題の有無に基づく「けど」事例の分類

		前件の発話意図タイプ			
		信念	ゴール	手法	属性値
後件の発話意図タイプ	信念				
	(有)	2			
	(無)				
	手法				
	(有)	10	4		
	(無)	6/5	14	1	1
	属性値				
	(有)	4			
	(無)				
	真偽値				
	(有)	2			
	(無)				
真偽値:OK					
(有)	1	1			
(無)					
真偽値:EQ					
(有)	1				
(無)	3				

※括弧内は発話命題の有無を、数字は事例数を表す。

表3 「けど」の役割という観点からの「けど」事例のクラスタリング

		前件の発話意図タイプ			
		信念	ゴール	手法	属性値
後件の発話意図タイプ	信念				
	(有)	a			
	(無)				
	手法				
	(有)	a	e		
	(無)	b/c	f	f	f
	属性値				
	(有)	a			
	(無)				
	真偽値				
	(有)	a			
	(無)				
真偽値:OK					
(有)	a	a			
(無)					
真偽値:EQ					
(有)	a				
(無)	d				

※括弧内は発話命題の有無を、a～fはクラスタを表す。

れない場合があったためである。

次に各組合せパターンに分類された「けど」事例の性質を調べ、「けど」の役割という観点から「けど」事例のクラスタリングを行った。すなわち、前件と後件に対する「けど」の役割が同じものを1つのクラスタにまとめた。この結果を表3に示す。表3において、同じアルファベットは同じクラスタであることを示している。

最後に各クラスタに対し、「けど」の役割に応じた意味解析のためのルールを設計した。以下、各クラスタにおける「けど」の役割と、そのクラスタに与えられた意味解析ルールについて述べる。なお、意味解析ルールはA→Bという形で表記されるが、Aは「けど」を解析する前の記述、Bは解析後の記述を示す。

#### クラスタ a (独立型)

このクラスタに属する「けど」事例では、「ロイへのメッセージはできましたけど保存して送ったらいいんですか」や「トゥーの所書き終ってシーシーに書きたいんですけどそのときはカーソルで動かしたらいいんでしょうか」のように、「けど」の前件でユーザの信念/ゴールが述べられ、後件でその信念/ゴールの下での発話がなされている。しかしながら、前件と後件のこのような関係は、

「けど」によって接続されなくても、すなわち前件と後件が2つの文として述べられたとしても、成立すべきものである。従って、このクラスタに属する「けど」事例に対しては、前件と後件を独立したものとして考え、「前件の発話意図記述、けど、後件の発話意図記述 → 前件の発話意図記述、後件の発話意図記述」という意味解析ルールを与えることができる。

以上で述べたように、このクラスタにおける「けど」の役割は弱い。しかしながら、その一方で、後件における照応解析/省略解析のための先行詞候補/補完候補が前件で与えられている場合も多い。例えば、最初の「けど」事例では、「保存する」や「送る」の対象格になる「ロイへのメッセージ」が前件で与えられているし、二番目の「けど」事例でも、「その」の先行詞になる行為「書く」が前件で与えられている。従って、このクラスタに属する「けど」事例の後件において、照応解析/省略解析を行う際には、前件で述べられた「もの」や「行為」、「状態」を照応解析/省略解析の候補として優先的に扱えばよい。

#### クラスタ b (説明I型)

このクラスタに属する「けど」事例の特徴は、前件で直前に起きた/生じた事象(出来事/状態)

について言及し、後件で発話命題なしの手法に関する質問を行っているという点にある。すなわち、ユーザは、前件で述べた事象を不正なものと考え、後件でその解消方法を尋ねている。実際、このクラスタに属する「けど」事例では、「メッセージが消えちゃったんですけどどうしたらいいんでしょうか」や「インボックスが消えてしまったんですけどどうしたらいいんですか」といった言い方がされている。従って、クラスタ b 用の意味解析ルールは「前件の発話意図記述、(手法 NIL) → 前件の発話意図記述、(手法 (発話命題 3))」となる。但し、(発話命題 3) は前件で述べられた事象が起きる/生じる以前の状態にするための行為フレームとして生成される。例えば、「メッセージが消えちゃったんです」という前件に対しては、現在の状態をメッセージが消える前の状態にする行為フレームが生成され、(発話命題 3) となる。もし、行為フレームを生成できない場合には、その旨がユーザに告げられる。

なお、次のクラスタ c に属する「けど」事例は、このクラスタには属さないものとする。

#### クラスタ c (説明 II 型)

このクラスタはクラスタ b の特殊な場合と考えられる。すなわち、このクラスタに属する「けど」事例は、クラスタ b の場合と発話意図タイプの組合せ方が同じであるが、前件の発話命題がある発話命題の否定という形の場合に限られる。従って、クラスタ c 用の意味解析ルールは「(信念 (発話命題 1) の否定)、けど、(手法 NIL) → (信念 (発話命題 1) の否定)、(手法 (発話命題 1))」となる。例えば、このクラスタに属する「けど」事例「メッセージが出てこないんですけどどうしたらいいんですか」は (信念 (メッセージが出てこない))、(手法 (メッセージが出てくる)) という 2 つの発話意図記述として処理される。

#### クラスタ d (説明 III 型)

このクラスタもクラスタ b の特殊な場合と考えることができる。クラスタ b に属する「けど」事例との違いは、後件の発話意図タイプが「真偽値:EQ」という点にある。この発話意図タイプの場合、3. で述べたように、対応する発話命題は状態フレームでなければならない。従って、このクラスタ用の意味解析ルールは「前件の発話意図記述、けど、(真偽値:EQ NIL) → 前件の発話意図記述、(真偽値:EQ (発話命題 3))」と設計される。但し、(発話命題 3) は前件で述べられた状態/前件で述べられた行為によって生じた状態を表現する状態フレームとして生成される。

このクラスタに属する「けど」事例は、例えば「インボックスに取り込んでしまったんですけど

よかったですか」や「1行空いてしまうけどいいんですか」などである。

#### クラスタ e (重複型)

このクラスタに属する「けど」事例の特徴は、前件の発話意図タイプが「ゴール」、後件の発話意図タイプが「手法」であるという点にある。このことが示すように、前件で述べた発話意図と関連した内容の発話意図を後件で重複して述べている。そこで、クラスタ e 用の意味解析ルールとして「(ゴール (発話命題 1))、けど、(手法 (発話命題 2)) → (ゴール (発話命題 3))」という形のもの設計する。但し、(発話命題 3) は以下のアルゴリズムに従って生成される。

もし、(発話命題 1) と (発話命題 2) が同定可能ならば、同定し、その結果を (発話命題 3) とする。同定できなければ、各発話命題を達成するためのプランに重複がないかどうかを調べ、上位プランに対応する発話命題を (発話命題 3) とする。もし、プラン間に重複関係がなければ、その複文発話文には何らかの誤解が含まれているものと考え、誤解の存在とその内容を提示するとともに前件の発話命題を (発話命題 3) として処理を続ける。

#### クラスタ f (強化型)

このクラスタに属する「けど」事例には、例えば「メッセージの読み方がわからないんですけどどうしたらいいんですか」や「ドラフト 2 を消したんですけどどうすればいいんですか」といったものがある。これらの例を見ても分かるように、複文発話文が有する発話意図は前件で述べられており、その発話意図をより明確にするための手段として後件が述べられている。しかしながら、このような発話意図の明確化は、人相手には効果的であるが、計算機相手に話すときは不必要である。従って、「前件の発話意図記述、けど、(手法 NIL) → 前件の発話意図記述」というルールが成立する。

## 5. 「けど」を含む複文の発話意図解析

「けど」を含む複文の発話意図解析は、図 1 に示された流れに沿って行われる。

### 5.1 形態素解析

入力された発話文は、まず JUMAN[11] によって形態素解析される。JUMAN は、入力発話文を適当な形態素に分解し、各形態素に原形や品詞名、品詞細分類名、活用形といった情報を付加す

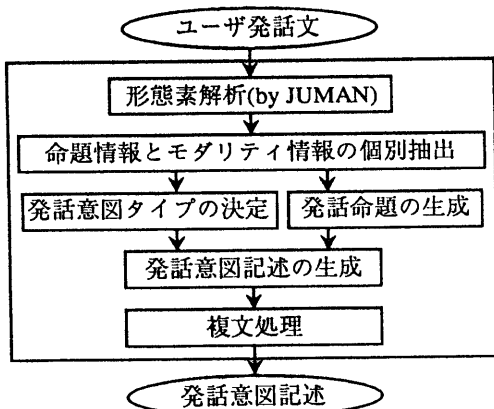


図1 発話意図解析の流れ

る。なお、本稿の文法用語は、JUMANで採用されている益岡・田窪文法 [3] に依っている。

### 5.2 命題情報とモダリティ情報の個別抽出

命題情報とモダリティ情報は、JUMANの出力である形態素列から個別に抽出される [10]。その様子を表4に示す。

モダリティ情報には、意味語や品詞情報 (品詞名/品詞細分類名)、活用形情報、原形情報の4種類がある。意味語は、発話意図タイプの決定に影響を与える意味情報の種類と定義される。例えば、話者の「良い」という判断を表す形容詞「いい、よい、よろしい、…」からは意味語「OK述語」がモダリティ情報として抽出される。

一方、命題情報には、原形情報、品詞情報、活用形情報の3種類があり、タスク領域上で定義可能な「もの」や「行為」、「状態」を表現する語 (名詞や動詞、形容詞など)、およびそれらの関係を表現する語 (助詞など) から抽出され、そうでない語からは抽出されない。また、判定詞や接尾辞、助動詞、助詞のような語からは、命題情報が抽出される語に接続した場合のみ、命題情報が抽出される。

### 5.3 発話意図タイプの決定

発話意図タイプは、その発話文から抽出されたモダリティ情報に発話意図タイプ決定ルールを適用することによって決められる。ここで文献 [9] で得られた発話意図タイプ決定ルールの一部を、例として表5に示す。また、ルールの記述形式を拡張バックス記法で示す。

(ルール) ::= ((パターン部) (発話意図タイプ))  
 (パターン部) ::= ((パターン) (パターン)\*)  
 (パターン) ::= ((特徴量) (特徴量)\*)

表4 命題情報とモダリティ情報の個別抽出の例

モダリティ情報	対応する入力形態素	命題情報
普通名詞	メッセージ	メッセージ 普通名詞
	が	が 格助詞
状態述語	出てこない	出てくる 動詞 未然形 ない 形容詞性述語接尾辞 基本形
	んです	
	けど	けど 述語接続助詞
HOW句	どうしたら	
OK述語	いい	
	んです	
疑問助詞	か	
文末		

表5 発話意図タイプ決定ルール (一部)

((疑問語)(OK述語))	属性値:OK)
((疑問語)(基本形 文末))	属性値:OK)
((疑問語))	属性値)

(特徴量) ::= (意味語) | (品詞情報)

| (活用形情報) | (原形情報)

次に文献 [9] で提案したルール適用法を示す。

【ルール  $r$  ( $1 \leq r \leq R$ ) の適用】モダリティ情報にルール1からルールRを順次適用する。モダリティ情報とルール  $r$  のパターン部とのマッチングに成功したら、対応する発話意図タイプを出力し、ルールの適用を終了する。

【パターン部のマッチング】パターン部の最左パターンから最右パターンへとマッチングを行い、全パターン部のマッチングに成功したら、そのパターン部のマッチングは成功である。

【パターンのマッチング】当該パターンがモダリティ情報中のある部分列と同一であるならば、そのパターンのマッチングは成功である。成功したら、その部分列よりも文末方向にある特徴量の列を新たなモダリティ情報とし、次のパターン部のマッチングを行う。

このルール適用法は単文相当の発話文 (単文もしくはせいぜい1つの遂行動詞を有する複文) [10] を対象としている。そこで、「けど」を含む複文を取り扱えるよう拡張する必要がある。表4に示されているように、モダリティ情報に複文解析のトリガーとなる「けど」が含まれている場合、そのモダリティ情報を「けど」の前後で分割し、「普通名詞 状態述語 文末」および「HOW句 OK述語 疑問助詞 文末」という2つのモダリティ情

報を得る。このとき、前件のモダリティ情報には意味語「文末」が加えられた。そして、各モダリティ情報から、文献[9]の手法に基づいて、それぞれの発話意図タイプが決定される。結局、表4にあるような発話の前件からは「信念」が、後件からは「手法」が発話意図タイプとして得られる。

#### 5.4 発話命題の生成

発話命題は、抽出された命題情報を意味主導で解析することにより、生成される。命題情報に複文解析のトリガーとなる「けど 述語接続助詞」が含まれている場合、その命題情報をこのトリガーの前後で分割する。表4の例では、命題情報は「メッセージ 普通名詞 が 格助詞 出てくる 動詞 未然形 ない 形容詞性述語接尾辞 基本形」と「」に分割される。但し、命題情報「」という表記は、文献[10]で示した、命題情報が抽出されなかった状態を示すものとする。結局、表4のような発話文からは「メッセージが出てこない」を表現する状態フレームと、命題情報なしを意味する記号NILがそれぞれの発話命題として次ステージに渡される。

#### 5.5 発話意図記述の生成

「けど」事例からは、「けど」の前件と後件からそれぞれ発話意図タイプと発話命題が生成された。このステージでは、対応する発話意図タイプと発話命題を組み合わせるにより、2つの発話意図記述が決定/生成される。次ステージにはこの2つの発話意図記述と発話意図記述間の関係を表す「けど」ラベルが渡される。

#### 5.6 複文処理

前ステージから渡された情報の中に「けど」ラベルが含まれているとき、以下の処理が実行される。

2つの発話意図記述の発話意図タイプの組合せパターンと後件の発話命題の有無を調べ、表3に基づいて、対応する意味解析ルールを適用する。表4の例の場合は、クラスター用の意味解析ルールが適用される。その結果、前件の発話意図記述「(信念 (メッセージが出てこない))」と後件の発話意図記述「(手法 NIL)」は、変換され、「(手法 (メッセージが出てくる))」として処理される。

## 6. まとめ

本稿では、実際の支援対話で用いられた述語接続助詞「けど」を分析し、「けど」の前件と後件の関係を発話意図タイプの組合せパターンおよび後件の発話命題の有無に基づいて分類した。そして、この分類に基づいて、「けど」の役割を計算機

処理可能な形式で記述した。具体的には、それぞれの役割に対し、意味解析ルールを設計した。

本稿で対象とした複文は、述語接続助詞「けど」によって接続されたものに限られた。「けど」は、文法書などによれば、逆接を表す接続助詞とされているが、実際の対話場面ではより多様な意味で用いられている。このような言葉はほかにも多く、文法書や辞書の類から話し言葉理解のための意味辞書/概念辞書を作成するには無理がある。そこで我々は、個別の言語現象を1つずつ解析するというアプローチを採用している。

観察した事例が少ないため、表2/表3には空白欄がある。このようなスパース性を解消することは今後の課題となる。

## 参考文献

- [1] 熊本忠彦, 伊藤昭, 海老名毅: “電子メール利用支援システム,” 郵政省通信総合研究所第87回研究発表会予稿, pp. 36-43 (Nov. 1994).
- [2] 新村出(編): “広辞苑”, 第四版, 岩波書店, 東京(1991).
- [3] 益岡隆志, 田窪行則: “基礎日本語文法”, くろしお出版, 東京(1991).
- [4] 森岡健二, 徳川宗賢, 川端善明, 中村明, 星野晃一(編): “集英社国語辞典”, 第一版, 集英社, 東京(1993).
- [5] 金田一京助, 見坊豪紀, 金田一春彦, 柴田武, 山田忠雄(編): “新明解国語辞典”, 第三版, 三省堂, 東京(1981).
- [6] Peek, J.: “Mh and Xmh - E-mail for Users and Programmers -,” O'Reilly & Associates, Inc. (1991).
- [7] Kumamoto, T., Ito, A. and Ebina, T.: “Design and Construction of an Advisory Dialogue Database,” IEICE Trans. Information and Systems, E78-D, 4, pp. 420-427 (1995).
- [8] 益岡隆志: “モダリティの文法”, くろしお出版, 東京(1991).
- [9] 熊本忠彦, 伊藤昭, 海老名毅: “支援対話におけるユーザ発話意図の認識”, 信学論D-II, J77-D-II, 6, pp. 1114-1123 (Jun. 1994).
- [10] 熊本忠彦, 伊藤昭, 海老名毅: “命題情報とモダリティ情報の個別抽出に基づく発話意図解析手法について”, 人知研資, SIG-SLUD-9403-2, pp. 7-12 (Feb. 1995).
- [11] 妙木裕, 松本裕治, 長尾真: “汎用日本語辞書および形態素解析システム”, 第42回情処全大, 1C-9, p. 3-17 (1991).