

論文 / 著書情報
Article / Book Information

論題(和文)	セミナーでの対話にみられるパラフレーズの分析
Title(English)	Analysis of Paraphrases and Cohesion in Dialogues at Seminars
著者(和文)	仁科喜久子
Authors(English)	KIKUKO NISHINA
出典(和文)	文部省科学研究費補助金重点領域「音声・対話・概念の対話理解と生成に関する研究」研究成果報告書, Vol. , No. , pp. 59-62
Citation(English)	Research on Understanding and Generating Dialogue by Integrated Processing of Speech, Language, Concept Report of Grant-in-Aid for Scientific Research by Ministry of Education, Science and Culture, Vol. , No. , pp. 59-62
発行日 / Pub. date	1994,

セミナーでの対話にみられるパラフレーズの分析

仁科 喜久子

東京工業大学

1 本研究の目的

本研究は音声対話を語、文からディスコースに至るまで言語学的に分析し、そのアルゴリズムを構築することを目標とする。

ビデオ収録した自然な音声データに見られるさまざまな現象を特にパラフレーズを中心に分析し、将来の技術的な可能性に託して、言語学の立場から工学的な開発に協力できることを願うものである。

2 分析対象

従来の音声対話コーパスは工学的に扱うことが可能な範囲でのタスクによる模擬会話実験がほとんどであるが、あえて本研究は、現状では技術的にむずかしい自然で複雑な対話ディスコースを扱う。この目的のためには6月から9月の間に東工大大学院のセミナーで外国人留学生および日本人学生が発表するものを10本ビデオ録画した。

このデータからは(1)雑談ではない実用的な目的のある対話を観察でき、(2)話し手の知識背景(世界)による対話の様相の比較できると思われる。

さらにこの録画ビデオの書き起こしを行った。テープ起こしに際しては学生のハンドアウトや発表原稿のコピーを譲り受け、さらに分からないところは、本人に面接して訂正、確認をした。

3 分析方法

本稿ではノンバーバル要素を除き言語要素のみを分析対象とする。発話が文法的に文を成していると認定できる場合は「。」をつけた。文法的には文が完結していないか、発話者がまだ言い終わっていないうちに、次の発話者が発話する場合「//」

を文あるいは句末につけた。

資料を分析する方法として語、構文、ディスコースの各レベルで考えられる。本稿ではこれらの各レベルでのパラフレーズに注目する。

4 対話中のパラフレーズの定義

図1はこのような状況における対話の構造をモデル化したものである。

(1) Xは概念、Sは話し手、Uはある話し手から出た発話である。ある発話の場では有限の話し手と有限の発話がある。つまり一定の時間内の実際の対話を採録すると、話し手が n 人で、発話の回数は m 回と数えられる。対話の定義としては「話し手、聞き手が一対一で、両者の立場が随時交替するもの」とある。(「国語学研究事典」(明治書院)の「対話」の項目)ここでは、談話の場に複数の話し手、聞き手が存在しても、その発話と次の発話ではそれぞれ一対一となるという解釈で、 n 人による対話が成立するものとする。

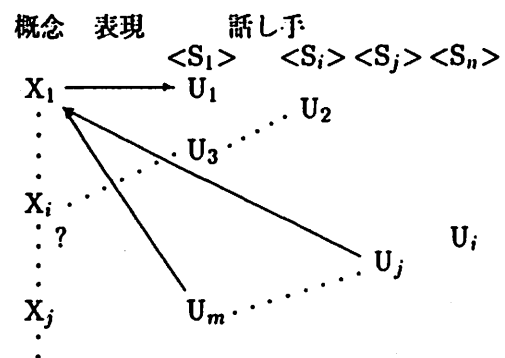


図1: 対話のパラフレーズ構造

(2) 一方、概念については発話の場に参加する話し手あるいは聞き手にとって、そこで言及される概念の数はおそらく有限ではないであろうが、それを確かめる方法はない。ここで下の図の意味を説明しよう。例えば話し手 S_1 は概念 X_1 を U_1 で

表現するとする。話し手 S_2 はこの表現を聞いて、いくつかの概念を対照してみるが、明確にある概念に特定できないことがある。そこで S_1 に対して発話 U_2 を質問として行う。 S_1 がその答 U_3 を発話するが、それでも明確でないときは、パラフレーズをして、 S_2 が別の表現で U_4 の質問をする場合もある。それに対してはじめての話し手が U_5 の答えを発話をすることもある。あるいは S_3 なる第3者が S_1 あるいは S_2 に対する対話者になることもある。

(3) 一つ概念が相互に理解された時点で、一連のパラフレーズの対話は完結すると考えられる。別の見方をすれば、その談話の場において、対話の参加者が相互に理解したと見なされるところを談話の一つの単位とすることが出来る。理解したと見なされることを示すマーカー (K) として「分かった」「なるほど」「そういうわけですか」「はい」などの発話が予想される。これには客観的な判断法が必要である。ある場合はあいづちやノンバーバルな表現によることもあろうが、現在比較的把握しやすいのは発話に現れるキーワードを基準にすることである。将来はこれらのキーワードやノンバーバルサインの統計的処理による単位の認定も可能であろう。しかし現在のところは分析者の判断によるものとする。

このような概念と表現の相互理解の構造は、語彙レベルから構文、ディスコースのレベルで記述することができる。

(4) 言語単位によるパラフレーズの分類

語、文、ディスコースのそれぞれのレベルでパラフレーズが起こるものを、次のような式で示す。

1. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_i(x)(w_1 \cdots w_h)$
言い淀み、つぶやき、説明の明確化
2. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_i(x)(S_1 \cdots S_h)$
説明、確認 (+ つまり、すなわち)
3. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_i(x)(S_1 \cdots S_h)$
説明、言い淀み、つぶやき、自省
(+ つまり、すなわち)
4. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_i(x)(w_1 \cdots w_h)$
強調、確認 (+ つまり、すなわち)
5. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_j(y)(w_1 \cdots w_h)$
模倣、つぶやき (自省)、確認、疑問
6. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_j(y)(S_1 \cdots S_h)$
確認、疑問

7. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_j(y)(S_1 \cdots S_h)$
模倣、つぶやき (自省)、確認、疑問
8. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_j(y)(w_1 \cdots w_h)$
確認、疑問
9. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_k(x)(w_1 \cdots w_h)$
説明、確認、強調 (+ つまり、すなわち、と)
10. $U_i(x)(w_1 \cdots w_l) \xrightarrow{P} U_k(x)(S_1 \cdots S_h)$
説明、確認、強調 (+ つまり、すなわち、～と)
11. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_k(x)(S_1 \cdots S_h)$
前言の説明、確認、強調、要約
(+ つまり、すなわち、と)
12. $U_i(x)(S_1 \cdots S_l) \xrightarrow{P} U_k(x)(w_1 \cdots w_h)$
前言の要約、確認、強調

ある対話の中で i 番目の発話 (U) において話し手 S が語 w_1 を発話したものを、もう一度同じ語 w_1 と発話するとき、パラフレーズが起こったとする。

1. は同じ発話者の同じ発話位置で $w_1 \cdots w_l$ の語の連なりが、全く同じか、変形された形で、語の連なりが発話される場合である。語の連なりは $0 \leq l, k$ であり、1語以上の語が1語以上の語にパラフレーズされる。左辺と右辺の語の連なりの数は同じでなくてもよい。

2. は1. と発話支者と発話位置は同様であるが、始めの発話で語のものが、文でパラフレーズされる場合を示す。

3. は1., 2. と同じように同一発話者と発話位置内であるが、始めの発話が文であり、パラフレーズも文になる例である。

4. はその逆でもとの表現が文であるものが、語の連なりになる例である。語の数は1でもよい。

5. から8. は、発話者 X が発話 U_i をした後、別の話者 y がする発話 U_j の中でパラフレーズが起こる場合のそれぞれの語、文の組み合わせによるケースである。このときの j は $i < j$ であり、 i の直後 ($j = 1+i$) であることが多い。

9. ~ 12. は、5. から8. のようなケースが U_i とおなじ発話者であるというケースである。すなわち、話者 (S) が対話の流れの中で、発話の位置が1以上スキップする場合である。 ($k \leq i+1$)

5 語彙レベルからみたパラフレーズ

語彙レベルでは発話の中でパラフレーズされる語句の関係が概念シソーラス上でどうマッピングされるかを分析し記述する。また頻度を調べ、その傾向を考察し、いくつかのパターンに分類する。

今回の分析では、日本語能力の高くない留学生との対話中で、概念がわかっているが、日本語としての専門的、慣習的表現あるいは一般的な表現を知らないために教官あるいは日本人学生による訂正作業としての他者間のパラフレーズが多くみられた。日本人あるいは日本語能力の高い留学生の間では、前者のような事例は少ない。思考内容を最も適切に表現したいために、自己完結的なパラフレーズや言い淀みがパラフレーズとしてあらわれる。

6 文レベルのパラフレーズ

ある陳述文がその後どのような形でパラフレーズされ、その文の要素は構文論的にどのようなプロセスで変形しているかを観察する。

$$S_1 + S_2 \rightarrow S_3 = S'_1 + S'_2$$

概念 → 表現

$$X \rightarrow U(S = N \cdot Pr)$$

例2: 話し手 <A> 学生、 教授

A: デジタル信号を応用するリップルは望ましくない場合があります。そのとき帯域のリップルはチクセキします。クライを落とします。このときには単調な周波数特性は必要です。

B: チクセキする、クライを落とすってどういう意味ですか。

A: このリップルはチクセキすればあまりよくないという意味です。

B: 特性が劣化するという意味、そういうことね //

A: はい //

B: degrade //

A: 単調な周波数特性が必要です。

B: 特性が劣化すること。リップルが蓄積して特性が劣化する。ということね。

A: 特性が劣化します。

7 ディスコースレベルからみたパラフレーズ

ディスコースでは同一の話し手のパラフレーズ、対話者間のやりとりによるパラフレーズなどを文を越えて起こる場合も扱う。つまりディスコースのコヒージョンの構造を記述することになる。こ

のレベルでは、うまく伝わらない場合の質疑応答のプロセスを追って行く中で、そこで用いられる指示語、接続語、言及内容を括る述語に注目する。特にセミナーにおける対話では複数の対話者間で理解に至るための発話が見られ、それを示す種々のいわばマーカがこれらの語句の中に見られる。

また、対話の流れの中ではその対話の行われる場の背景、さらには発話者の担う文化と言語能力の影響を考慮しなければならない。セミナーの対話では社会位相的な発話の配慮はあまりないという予測でも分析を始めたが、今回のデータでは複数の発話者間で行われるのが見られた。

ディスコースを構造的に見るにはコヒージョンの記述が必要となるが、これを分析する方法としては対話中に現れる語句のコロケーションを統計的に調べる方法と理解に至るキーワードを抽出していく方法が有効だと思われる。

8 まとめ

(1) 理解のプロセスを示すキーワード

☆ 対話者の発話に対する理解を示す表現 = なるほど

☆ 旧情報の確認を示す表現 = ~わけですか。/っていうわけ.../っていうふうに理解したんですか。

☆ 質問内容の補充を示す表現 = その方法とこれと何が違うかっていうことですよ。僕が質問しているのは。

☆ 発話者自身の内容の明確化 = NPっていうか、NP...

☆ 発話者自身の内容の要約と念押し = 何言いたいかって言うと、要するに、

☆ 対話者の発話内容の要約と確認 = 要するに、~と、解釈しましたけど。~言いたいわけでしょう。

(2) コヒージョンを示す語句の検索

コロケーションを見る。(統計的方法)

キーワードを探す。= 要するに/つまり/となるわけです。

(3) 言語能力と専門知識によるパラフレーズの現れ方の傾向

パラフレーズの現れ方の数を、各資料毎に示すと、表1のようになる。日本語能力と専門知識を仮に点数で表現してみたが、日本語能力は日本人

資料番号		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
国籍		ブルガリア	タイ	イラン	日本	インド	台湾	タイ	フィリピン	日本	日本	日本
母語		ブルガリア	タイ	ペルシャ	日本	タミール	中国	タイ	タガログ	日本	日本	日本
学習歴		1年	1年	3年	--	6ヵ月	6ヵ月	1年	6ヵ月	--	--	--
滞日期間		2年	1年半	10年	25才	4年	6ヵ月	1年半	2年	23才	28才	19才
学年		D1	D1	D3	D1	D3	M1	M1	M1	M1	D2	B1 (3名)
主な対話相手 (教育)		N	N	N	N	T1	I	K	T2	T2	T1	Y
パラフレーズのレベル	語	5	4	12	8	18	82	12	27	10	9	14
	1頁平均	2.5	1	6	3.2	3.2	<u>12.6</u>	<u>12</u>	<u>10.8</u>	3.3	1.8	0.9
	文	10	23	13	6	42	25	0	10	11	31	63
	1頁平均	5	5.7	<u>6.5</u>	2.4	<u>7.6</u>	3.8	0	3.2	1.2	<u>6.2</u>	3.9
	ディスコース	2	4	5	4	9	11	2	8	9	5	3
収容量 (頁)		2	4	2	2.5	5.5	6.5	1	2.5	3	5	16
日本語能力 (点)		4	5	8	10	6	4	3	3	10	10	10
専門知識 (点)		8	8	10	8	10	6	6	6	6	9	2
日本語能力 + 専門知識 (点)		12	13	18	18	16	10	9	9	16	19	12

表1: パラフレーズの分類と傾向

学生を10点として、留学生は日本語学習期間、滞
在期間、教室のレベルでのから、専門知識は博士
3年を10点として、以後は各学年を降りる毎に-1
とした。

9 今後の課題

以上本稿では収録したデータの一部を例として
音声対話をディスコースの観点で考察し、そのプ
ロトコルを求めようとした。今回セミナーの対話
を分析するに当たっての予想としてはディスコー
ス単位の認定がしやすく、コピーレントな対話が
なされるだろうということであったが、ほぼ予想
通りの結果が観察できた。この考察を通して、対
話の定義、対話の単位、パラフレーズの範囲の認
定、文の完結性などさまざまな問題が出てきた。
しかしセミナーの対話の特色を知るためには、他
のさまざまなタイプの音声ディスコースと対照比
較しなければならぬ。他にも講義ビデオやトピッ
クだけ与えられた電話での自由対話など参照したが、
ここでは議論することができなかった。さらにテ
レビやラジオの対話も新たな資料として分析し、
視覚的補助の有無による違いなども分析する必要
がある。このような詳しい対照研究と収録したデー
タについては次の機会に発表したいと考えている。

- (a) 語のレベル = 言い替えや繰り返しのタイプ
別に分け、言い替えに関してはシソーラスと
の対照の頻度を含む位相マップを各ジャンル

ごとに作成してまとめとする。

- (b) 文レベル = 誤解の起こる場面あるいは起こ
ると推定する話し手の思考プロセスを各ジャン
ルごとにまとめ、構文的な特徴を記述し、
そのプロトコルを作成する。
- (c) 各データをコピージョンの観点から分析し、
アルゴリズムの構築を計る。
- (2) それぞれのジャンル・分野による異同を観察
する。
- (3) 音声対話のパラフレーズの形式化に関する研
究報告書を作成する。内容としては、分析対象の
収録ビデオとその書き起こしデータベースを含め
た分析対象資料と分析の結果を資料編として掲載
する。