

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とコヒージョンの分析-概念関係との対応-
Title(English)	Analysis of Paraphrases and Cohesion in Dialogues at Seminars-Relationships between concept and works-
著者(和文)	仁科喜久子, 笹川洋子, 土井みつる
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. 71-78
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. 71-78
発行日 / Pub. date	1995,

音声対話ディスクコースにおけるパラフレーズと
コヒージョンの分析とその形式化の研究
セミナーでの対話にみられるパラフレーズとコヒージョンの分析
-概念関係との対応-

仁科喜久子

(研究協力者; 笹川洋子・土井みつる)

1. 概要

本年度は前年度のセミナーの収録データに更に6本のデータを加えた。また分析は外国人および日本人学生のセミナー発表時に見られる教官その他の研究室メンバーとの対話の中で特にパラフレーズに注目して行った。

本報告ではディスクコースの枠組みの中で語のコロケーションを観察することで知識をベースにしたコヒーレンスとテキストに依存するコヒージョンの構造を考察し、ディスクコース中に出現する語句を概念マップに置き換える試みをする。また、それに加えてディスクコース・マーカーを観察することで発話時における曖昧性を解消する可能性を検討する。

2. 収録データ

昨年度は東京工業大学大学院理工学研究科セミナー60分×10回分を収録し、その中でパラフレーズ部分を抽出し、書き起こし資料完成を完成した。本年度分は60分×6回分を収録書き起こし中である。細目は以下の通りである。

資料1 <電気電子研究科大学院セミナーN研究室セミナー> (録画時間60分)

1993.6. 最大平坦型FIRヒルベルト変換器に関する研究

発表者K.S (博士2年生 国籍 ブルガリア)

資料2 <電気工学科 F研究室大学院セミナー> (録画時間60分) 1993.9.30

ACOUSTIC ECHO CANCELLATION USING SUBBAND STRUCTURE に関する研究

発表者D.D (博士課程1年生 国籍 タイ)

資料3 <電気工学科 N研究室大学院セミナー> (録画時間60分) 1993.9.13. 最大平坦

線形位相デジタル・フィルターの構成とその雑音特性の解析 発表者S.S (博士課程3年 国籍 イラン)

資料4 <電気工学科N研究室大学院セミナー> (録画時間30分) 1993.9 MOSトランジスタによる低電圧回路に関する研究

発表者N.S (博士課程1年生 国籍 日本)

資料5 <情報工学科T研究室セミナー> (録画時間60分) 1993.5.17 Theoretical Study on the Family of GLR Parsings Algorithms Using Ancestors Table

発表者K.G.S (博士課程3年 国籍 インド)

資料6 <経営工学科 大学院セミナー (輪講) I研究室> (録画時間60分)

1993.6. Influence of Experience on Age Differences in Cognitive Functioning.

発表者 C.Y (修士1年 国籍 台湾)

資料7 <化学工学科K研究室ゼミ> (録画時間60分)

1993.5. A Liquid Emulsion membrane Process for Separation of Amino Acid

発表者 W.W (修士課程1年生 国籍 タイ)

資料8 <金属工学科T研究室大学院セミナー: 雑誌会> (録画時間60分) 1994.9.30

Effect of Heat Treatment on the Co-oxidation Behavior of Amorphous Fe-Cr-P-C and Fe-Ni-Cr-P-B Alloys in 1N HCL

発表者 R.Y (修士課程1年生 国籍 フィリピン 日本語学習歴6カ月 滞日期間2年)

資料9 <金属工学科T研究室大学院セミナー> (録画時間60分) 1993.9.30. 非接触測定

法による有機塗膜のふくれのモニタリング 発表者S.S (修士過程1年生 国籍 日本)

資料10 <情報工学科T研究室大学院セミナー> (録画時間60分) 1993.6 左隅構文解析法

による優先回探索の基礎

発表者A.A (博士課程3年生 国籍 日本) (本年度採録分)

資料11 <原子核工学科T研究室大学院セミナー> (収録時間60分) 1994.11.4

Spectroelectrochemical Study of Uranium and Cerium in Carbonate Media

発表者 R.S (修士1年生 国籍 韓国)

資料12 <原子核工学科T研究室大学院セミナー> (収録時間60分)

1994.11.4 論文紹介

Laser-Induced Europium Luminescence

上記A, Bの場合にコミュニケーションブレークダウンが起こり、それを修復するひとつの言語的なストラテジーとして、繰り返しやパラフレーズが用いられる。

C:聞き手を説得し、念を押すためのレトリックとして用いる。

これはブレークダウンを起こすものではないが、上手な対話の中で見られる。

4. ディスコース中のパラフレーズ

上記で定義されるパラフレーズがコピージョン(結束構造, cohesion)の中でどういう機能をするかを検討する。

4.1 パラフレーズの類型とその例

パラフレーズの型としてレトリックの項目に準じて次の10種類に分けた。

- (a)置換(substitution)
- (b)省略(ellipsis)
- (c)照応(reference)
- (d)類義概念(synonymy)
- (e)上位下位概念(hyponymy)
- (f)換喩(metonymy)

換喩=本義と喩義の間に主として主体と属性の関係が見られるもの。

(g)提喩(synechdoche)=本義と喩義との間に部分と全体の関係があるもの。

(h)反意概念(antonymy)

例1<繰り返しの例>

K:ちょっと、ゆるめればって、どうゆるめるの？

A: A: は？

K: ゆるめればって、どうゆるめるわけ？

A: は？ (資料10)

例2 <類義>

I助教授: 要する、さらに、その、なんか経験とか、練習とか。

C: Y: うん。

I教授: トレーニングをつんだときの時の効果を、これでしょ。

I教授: だから、こー認知作業とか、認知処理やったときの

C: Y: うんうん、

I助教授: まあ性能、性能っていうとおかしいけど、能力とかね。そういうこと。

例3<上位下位概念>

C: Y. 321: 3番目、たぶん、あー専門、専門家についてのテスト、た、例えば、あのも、もともとは、うん、建築者、architect? architect? とか、 (資料6)

例4<換喩提喩>

K助教授114: んー。(21.2) えーっと、それ

から、(3.9)液膜の文献、が、たくさんあったって言ってたでしょ。 {途中省略}

K助教授146: (21.5) (・・・)でやく10個と伊藤さんが2つでしょ、それから、あそこの、もう1つ前の、あそこんところにもう1つあって、この間にイタリ一があったはずでしょ。

W. W 147: うん。

K助教授148: もう4つでしょ。あと6つあるんですか、6つぐらい。

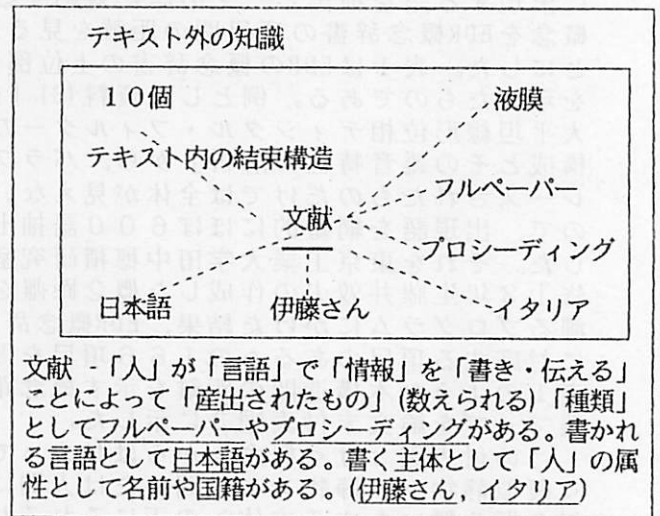
W. W 149: 日本語も、あります。 =

K助教授150: あ、日本語も。 (0.9)それは一、full paper ですか。(0.7)proceeding(・・・

W. W 151: ええ、そう思います。資料7

図2は、例4の語句についてのネットワークを示したものである。文献という概念が人名や国名などの固有名詞、文献内容、文献の種類の下位概念などと結びついている例であるが、このような事柄のフレームを登録することで、テキストの構造を記述するのである。

図2テキスト内の結束構造



文献 - 「人」が「言語」で「情報」を「書き・伝える」ことによって「産出されたもの」(数えられる)「種類」としてフルペーパーやプロシーディングがある。書かれる言語として日本語がある。書く主体として「人」の属性として名前や国籍がある。(伊藤さん、イタリア)

5. ディスコースマーカー

対話の流れを示すメタのマーカーとして接続の詞辞や指示語に注目することができる。対話の流れの接続(conjunction)構造すなわち結束性(coherence)がこれらの構文要素の機能の観察、分析からディスコースの構造の要素として記述できる。次にそれらの機能を6つに分けて示す。

(A)付加(additive)あと、それで、

- (B) 反意 (adversative) しかし, けれども,
- (C) 原因理由 (causal) だから, したら,
- (D) 時間推移 (temporal) それから, そして
- (E) 例示 たとえば,
- (F) 要約 要するに, つまり

これらの機能がどのように対話の中で用いられているかは、付表1に示した。

6. 対話のディスコースと概念関係

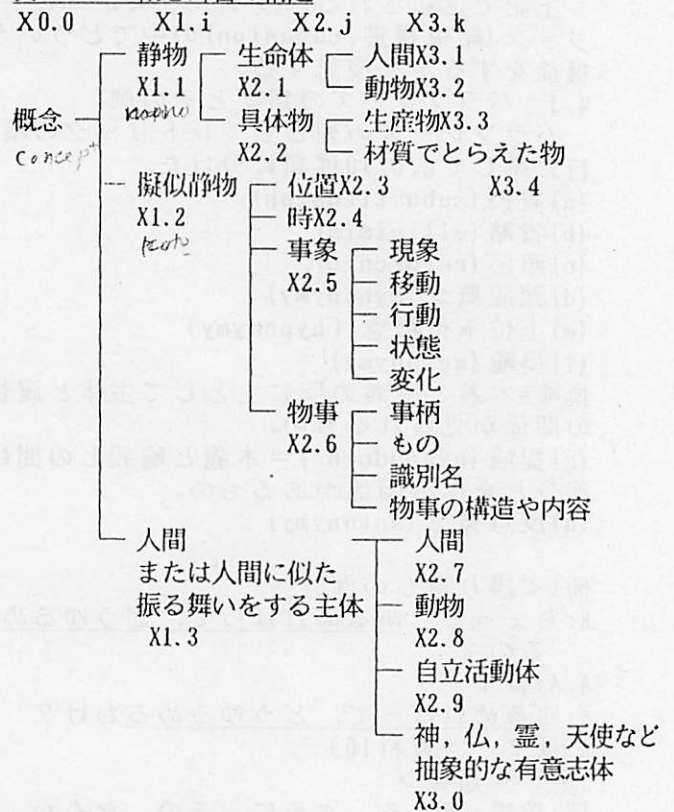
コロケーション (collocation・連語) とは周辺で起こる事実の上に成り立つ語と語との関係を扱うことである。パラフレーズは概念として関連する事柄を別の言語形式で表現することであるから、コロケーションと関連が深い。特に自然な状態で発話される対話音声認識する場合は、正確な構文構造をもっていないことが多い。そこでは文字表現された文章用の形態素解析、構文解析ツールが十分に機能することは期待しにくい。このような場合は、構文形式にこだわらずに、対話全体の流れをみて、そこに現れるコロケーション構造に着目するのも一方法であると思われる。

そこで、対話データに出現する語からコロケーションの様相を見るため対話データに出現する語を抽出し、その語に対応する概念をEDR概念辞書の項目間の距離を見ることにした。表1はEDRの概念辞書の上位部分を示したものである。例として資料(3)「最大平坦線形位相デジタル・フィルターの構成とその雑音特性の解析」から、パラフレーズされたものだけでは全体が見えないので、出現語を網羅的にほぼ600語抽出した。それを東京工業大学田中穂積研究室修士2年生藤井敦夫の作成した概念距離を測るプログラムにかけた結果、EDR概念辞書に対応する項目のあるもの166項目を得、表1のような木構造間の距離を示す概念距離マップを論文末付表図3に示した。

この付表図3は<概念>からはじまって<疑似静物><静物><人間または人間に似た振る舞いをする主体>の下にそれぞれの下位概念が連なり、データ中の出現語に対応する概念を示そうとするものであるが、紙面の都合上端末に至るまで図示しきれないため、省略した階層レベルの数を端末の概念の右下に示した。ここではマップ上に近いところに集まった概念項目はパラフレーズで表現される可能性の高い集合といえるものである。資料3のデータでは「増えるー増加する」「いるー必要」「せまいーせばめる」などがパラフレーズされている。

類義表現によるパラフレーズではネットワークの近いところのものが共起するが、比喩や提喩などではかなりかけ離れた概念との共起も有り得る。ただし、分量のあるディスコースでは、例えるものと例えられるものがそれぞれ多数出現し、そこでそれぞれの語群を成すこともあるはずである。

表1 EDR概念辞書の構造



1語に対して複数に対応する概念項目から最適の項目を選ぶとき今回は、筆者らの目で見て判断する方法をとった。これには客観的な説得力のある手続きが必要であるはずだが、本稿の目的は、それが可能になった先の問題について考察したいので、あえて主観の入る判断に頼った。

7. まとめ

7.1 分析結果

1) 口語の音声対話の認識をする上で、テキスト上の結束構造と知識との関連での結束性を区別する必要がある。データに現れる知識上の語の関連性は、概念シラを整備することで、発話の曖昧性を排除できる。ここではEDRの概念辞書を用いて、ディスコ

ース内に現れる語の概念関係の距離を調べた。一方、テキストに依存する結束性は、そのテキストの意味ネットワークを作成し、表現意図を絞ることで、表現形の決定が可能になる。

2) 対話には雑談、戦略的、論理構築などその目的によってさまざまな構造が考えられる。本稿で扱うセミナー対話については理想的な流れとして、対話者が相互に持つ背景知識を照応しながら、既知の知識から対話者の発話意図を理解し、新しい知見を両者が獲得していくことである。

3) 対話音声の性質にふさわしい文章テキストと違ったディスコース解析法の確立をする必要がある。それには、従来までの構文に強く依存しないコロケーション分析のアプローチも有効であると思われる。

7.2 今後の課題

1) セミナー対話データの収集（データベースへの登録）

2) コロケーションに関する精密な分析法の確立（特に対話音声の特色との関連）

3) 概念辞書に関する課題

a. EDR概念辞書を利用する場合、ディスコース中に現れる概念候補を一意に決定する方法を今後の課題としなければならない。

b. 上位下位関係の記述

c. 横の関係＝レトリックの構造（比喩、類喩）の記述

4) テキスト内概念・意味ネットワーク作成

5) ディスコースマーカ―の機能記述法の整備充実

6) ディスコース構造プロトコルの作成
知識と言語表現との照応関係の記述（類推過程の構造の記述）

参考文献

1) K. Ajimer & B. Altenberg (ed.) 1991
Corpus Linguistics Cambridge U.P.

付表

<参考資料1>

situation 2 (開始13分後)

S.S8: で、でーやはりそバングでのシュミレーションの結果で、えーノイズが低いっていうのが、ま、この式でも確かめられる、ということがわかります。

N教授9: えーとなんで低い。

S.S10: えーっと、理由はですね えー 理由はですね、このへん、ここから、えーっと、ここへ伝達関数っていうのが、えーっと かたちとしましては、えーっと、ま、このように、えーっと filterでいうと、パスバンド型filterになります。

N教授11: うん。

S.S12: ここからえーっと伝達関数は、＝

N教授13: うん。

2) M.A.K. Halliday 1976 Cohesion in English Longman

3) R. Baugrande, & W.D. Dressler 1981
Introduction to Text Linguistics
Longman (池上嘉彦他訳 1984 テキスト言語学入門 紀ノ国屋書店)

4) 仁科喜久子 1994 日本語表現習得の観点による概念体系の評価 日本電子化辞書研究所委託研究報告書

5) 仁科喜久子・笹川洋子・土井みつるほか 1994 理工系セミナーでの対話理解過程の分析ー理工系学生のシラバス作成に向けてー 日本語教育84号

6) 奥村学 1994 音声対話理解のための言語解析手法に関する基礎的研究ー語彙的結束性を考慮した漸進的語義曖昧性解消科研費重点領域 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研究成果報告書

7) Renkma, J. 1993 Discourse studies: an Introductory textbook John Benjamins

8) F. ロボ, 津田葵, 楠瀬淳三 1987
英語コミュニケーション論 大修館書店

9) 鶴丸弘昭 1994 対話文理解のためのソーラスの構築と対話文の解析

科研費重点領域 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研究成果報告書

10) D. Schiffrin 1987 Discourse Markers (Studies in International Socio-linguistics) Cambridge U.P.

11) D.A. Willkins 1976 Notional Syllabuses Oxford U.P.

(D.A. ウィルキンズ 島岡丘訳注 1984 ノーショナル・シラバス 桐原書店)

12) 山本幹雄・中川聖一 1992 文生成における同一概念に対する表現の一選択方法
人工知能学会全国大会

S. S14 : で、えーそれ、にえーつとその一、原因が非常に低いみたいなんですね。

N教授15 : ふーん。

S. S16 : そしてそのノイズが、えーつと 減衰、を受ける。

N教授 17 : どちらから。はーん。

S. S18 : たぶん、その、モジュールで発生するノイズが、えーつと出力に現れるまで非常に、えーつと、減衰この減衰が大きくて、ま全部として合わせても、それほどの、えーつと要素にはならない、ということができません。

N教授 19 : 信号一とノイズのパスと一緒にだからさ、信号も減衰される。

S. S 20 : そうです。けれどもノイズは非常にあれが小さいので、信号に相対的に(……) {音声減少}。

N教授21 : え？、信号が、あまり増幅されてないわけ？

S. S22 : そういうことも考えられます。

N教授23 : あっ、増幅、要するに信号のレベルがあ、あまり変わらずに伝わっていく。

S. S 24 : ええ。えーつと、これは、えーつと前にもN先生とお話ししていたんですけども、この、メ、バースタインの公式、っていうのが=

N教授25 : うん。

S. S26 : あの一、関数を近似するための、ま多項式型ですよね。関数多項式型。

N教授 27 : んーんー。

S. S 28 : それで一、えー、収束非常に遅いわけですよ。で実数を、どんどんあげていっても、あんまり、よく収束してくれないんですよ。

N教授 29 : うん。

S. S 30 : で、それが逆にこの場合は都合がいいんですよ。なぜかっていうと、こことここでの信号レベルがあまり変わらない。

N教授 31 : うん。

S. S32 : 信号レベルがあまり変わらないから、

N教授33 : うん。

S. S34 : あまり増幅されていないんですよ。発生するノイズが。

N教授35 : うん。 うん。

S. S 36 : (2.2)とその一隣合うモジュール同士の、

N教授 37 : うん。

S. S38 : えーつと、出力信号を比べてみると、

N教授39 : うん。

S. S40 : あまり大きな変化はない。

N教授41 : うーん。 =

S. S42 : ちょっとずつ、あの一たとえばローパスフィルターだったら、ちょっとずつその、スムーススムージングが行なわれていって一、

N教授 43 : うん。

S. S44 : いきなり、信号変わるのがもう、その、モデル、モデルっていうか、部分がないわけですよ。

N教授45 : うん。

S. S46 : 非常に(……)で収束が遅いので一、でそれが一(1.6)このような、そのノイズの低さにも、起因しているんですよ。

N教授47 : んー。じゃ、なるべくな、みんな、やって。 =

S. S48 : みんな、やって、

N教授49 : なにもしない =

S. S50 : なにもしない、ほとんどなにもしていないんです、1個ずつは。

N教授51 : はーん。

S. S52 : 1個ずつはコサインをかけるような感じ、

N教授53 : んー。

S. S54 : で一、それで一、みんな、力を合わせると、

N教授55 : んー。

S. S56 : けっこう、

N教授57 : いいかもしれない。 =

S. S58 : いいかもしれない。

表2

発話の機能とディスコース・マーカ

資料3:

●教官の発話 ○学生の発話 ↑上昇イントネーション

情報提供JT	情報要求JY ^a	確認要求KY	受信表示JH	確認要求応答KYO
(理解STAGE 1)				
S.S 8: OJT	やはり、……の結果で、……ということがわかります (根拠)			
N教授 9:	●JY なんで……↑			
S.S 10: OJT	理由はですねえ (理由) ……っていうのが、……としましては……でいうと (視点の限定)、……になります			
N教授 11:			●+JH うん	
S.S 12: OJT	ここから……は (根拠)			
N教授 13:			●+JH うん	
S.S 14: OJT	で、原因が ……みたいなんですわ (原因の推測)			
N教授 15:			●+JH ふうん	
S.S 16: OJT	そして……			
(理解STAGE 2)				
N教授 17:	●JY どちらから↑		●+JH はーん	
S.S 18: OJT	たぶん ……でも、それほどの……にはならないということが出来ます (条件の付加)。			
N教授 19:	●KY ……			
S.S 20: OJT	けれども…… (対比)			OKYO そうです
N教授 21:		●KY ……わけ↑	●-JH えっ↑	
S.S 22:				OKYO そういうことも考えられます
N教授 23:	●JT あっ……			
(理解STAGE 3)				
S.S 24: OJT	前にもN先生とお話していたんですけど……っていうのが (既知情報提示) ○-JH ええ			
N教授 25:			●+JH うん	
S.S 26: OJT	……	OKY ……ための……ですよ↑		
N教授 27:				●KYO んーんー
S.S 28: OJT	それで、……わけですよ (理由)。で、……でも……ないんですよ。			
N教授 29:			●+JH うん	
(理解STAGE 4)				
S.S 30: OJT	で、それが逆に都合がいいんですよ。なぜかっていうと、…… (理由)。			
N教授 31:			●+JH うん	
S.S 32: OJT	……から (理由)			
N教授 33:			●+JH うん	
S.S 34: OJT	……んですよ ……が			
N教授 35:			●+JH うん	
S.S 36: OJT	と、……の			
N教授 37:			●+JH うん	
S.S 38: OJT	……を比べてみると (比較)、			
N教授 39:				
S.S 40: OJT	……変化はない (変化)			
N教授 41:			●+JH うん	
S.S 42: OJT	たとえば (例示)、……だったら、……ていつて (変化の過程)			
N教授 43:			●+JH うん	
S.S 44: OJT	……っていうのか、……わけですよ (根拠)。			
N教授 45:			●+JH うん	
S.S 46: OJT	……ので、それが……にも 起因しているんですよ (原因)。			
(理解STAGE 5)				
N教授 47:	●JT じゃあ (前首の理解表示) なるべく (n°アリーズ1: みんなでやって) (*先取り発話)			
S.S 48: OJT	(n°アリーズ1: みんなでやって)		●+JH んー	
N教授 49:	●JT (n°アリーズ2: なにもしない) (*先取り発話)			
S.S 50: OJT	(n°アリーズ2: なにもしない) (n°アリーズ3: ほとんどなにもしてないんです) ……			
N教授 51:			●+JH はーん	
S.S 52: OJT	……ような感じ (例示)			
N教授 53:	んー			
S.S 54: OJT	それで ……と (理由)			
N教授 55:	んー			
S.S 56: OJT	(けっこう)			
N教授 57:	●JT (n°アリーズ3: いいかもしれない) (*先取り発話)			
S.S 58: OJT	(n°アリーズ3: いいかもしれない)			

図3 対話ディスコースの概念マップ

