

日本語の自由会話における談話構造の推定 ～因果関係の場合について～

西澤 信一郎 中川 裕志

横浜国立大学 工学部 電子情報工学科

E-Mail: {shin, nakagawa}@naklab.dnj.ynu.ac.jp

概要: 本稿では、日本語の会話中において、発話間の因果関係がどのような形式で記述されているのか、をコーパスを用いて検討した結果について述べる。このような談話構造は、発話者の「思考の流れ」を示しているものと考えられ、発話者は、地図課題対話など目的の定まった会話の場合はもちろんのこと、雑談など特定の目的に左右されない自由会話の場合でも、この構造をある程度認識し、協調的な会話を進めているものと考えられる。そこで、本稿では、地図課題など目的の定まった会話からなるコーパスではなく、飲み会の席上での会話データを対象とした自由課題コーパスを用いた検討を行なった。また、この検討結果を利用し、因果関係を記述するような談話構造をコーパス中から取り出すために必要な手順について提案した。

A Method of Discourse Structure Understanding for Causality Relation in Japanese Task-Free Dialogue

Shin'ichiro Nishizawa Hiroshi Nakagawa

Yokohama National University

E-Mail: {shin, nakagawa}@naklab.dnj.ynu.ac.jp

Abstract: We discuss here how a discourse structure representing a causality relation among two or more sentences is described in Japanese task-free dialogue. The structure, we think, shows “the stream of consideration” of a speaker. We argue that the speaker recognizes the structure and have a conversation cooperatively even in the case of task-free dialogue. Then, we discuss the causality relation with a corpus of Japanese task-free dialogue and show how to find out discourse structures for the relation from the corpus systematically.

1 はじめに

談話理解を考える際、いわゆる談話構造の認識は重要な要素となる。最も基本的な談話構造は、二つの発話間の接続関係であり、この関係、つまり発話対を認識することが、談話理解の基本となると考えられる。これについては、マニュアルや論文などのいわゆる書き言葉について、接続詞や指示語などによる接続パターンを用いてテキストの構造解析を行なう手法[福本 92, 田中 92]や、対話中の質問-応答を表す発話対の認識に関する研究[高野 93]などがある。これに対して本研究では、日本語による特に課題を設定していない状況での会話(自由会話)において、意味的な因果関係が「から」や「だって」などの接続(助)詞によって表されるような発話対を対象とし、(1)発話対の構造、(2)発話対を会話全体から発見(推定)するために必要な手順、について検討する。つまり、本稿では、[Hob90]にあるような、文の意味内容に関する接続関係を問題とするが、発話対が同一人物により発話されている場合だけでなく、それぞれの発話が別の話者によりなされている場合も検討の対象とする。

なお、本稿では、会話データとして図 1 のようなコーパスを用いる。

会話 24

1. O → P あのね、これでもいいんじゃない \
2. P → O わかった \
3. O → P えー、嘘で言ったんだよ \
4. E → O 何 /
5. O → E だって、牛乳入れろって言ってたらさー \
6. G → G 何か酒飲みたいなー \
7. K → G あっ、ありますよ \
8. G → K それ何 /
9. E → G モルツ \
10. P → G ウイスキー \
11. E → G うまいよ \

図 1: コーパスの例

このコーパスは大学のあるサークルでの飲み会の席上で録音された雑談(課題を特に設定していない自由会話)を、そこに同席した者がテキストに書き起こしたものであり、全部で 1980 の発話からなっている¹。また、このコーパスでは、因果関係を記述する語とし

¹書き起こす際に、(1): 発話の切れ目の認識、(2): 会話内容による

て、接続詞「だから / だって」および接続助詞(相当)「から / のだから / のだもの」が用いられている。

2 因果関係を表す発話対のとり構造

まず、因果関係を述べるような発話対としてどのような場合があるのかについて検討する。この時、最も基本的な発話対であると考えられるのは、因果関係の前件(原因 / 理由)および後件(結果)を述べている発話が隣接しているような場合である。ある発話が前の発話を受ける形でなされることにより、増進的に会話が行なわれることや、聞き手の理解のしやすさなどから考えて、このような発話対が会話中に現れる度合は非常に高いのではないかと考えられる。この時、発話対が同一人物による場合だけでなく、二つの発話が異なる話者による場合、つまり他人の発話を補足するような形で因果関係が述べられる場合もあると考えられる。

では、発話対が隣接発話ではないような場合はどうであろうか。これはつまり、ある発話 A と発話 B とが発話対をなす時、その間に他の発話が存在するということである。隣接する発話が発話対をなす場合が基本であるとする、ここで述べているように隣接していない発話同士が発話対をなす、ということ認識するためには、何らかの手がかりが必要となる。その一つとして、「話者の同一性」が用いられるのではないかと考えられる。すなわち、隣接していない発話が発話対をなす時、二つの発話が同一人物によって発話されていることが必要とされるのではないかということである。

以上の考察は、次のようにまとめることができる。

原則 1 発話の意味内容が因果関係により関係づけられる発話対は、次のような性質を持つ。

1. 基本的には、隣接する二発話が発話対をなす。さらに、二発話が同一人物により発話される場合が、聞き手による認識が最も容易であると考えられる。ただし、話者が異なる場合についても、発

るセグメント分け、(3): 話し手と聞き手のデータ追加、(4): 発話の末尾の調子のデータ追加、を行なった。例えば、「O → P あのね、これでもいいんじゃない \」という発話では、話し手が「O」で聞き手が「P」であり、末尾が下がり調子の発話であったことを示している。

話が隣接していることにより発話対として認識可能である。

2. 発話対が隣接しない二発話による場合、それぞれの発話は同一人物によって発話されていることにより、発話対として認識されると考えられる。

また、因果関係を述べる発話対の中で、接続(助)詞がどのような位置を占めるのかについては、すでに[西澤 96]などで、発話対が同一人物の発話による場合について述べられており、異なる話者による発話対の場合にも、これが当てはまると考えられる。つまり、次のようなことがいえる。

- 接続詞が用いられた場合、その性質から、(a): 「前件. 接続詞-後件」あるいは(b): 「後件. 接続詞-前件」、つまり接続詞を含む発話とその前方に位置する発話とが因果関係をなす、という構造が基本となる[三上 72, 森田 80].
- 上記に対する例外が認められる条件としては、「発話が“前件(原因/理由)-後件(結果)”という、前件と後件の生起の時間順序に従っている」ことが要求される。つまり、(b'): 「接続詞-前件. 後件」は談話構造として認められる可能性があるが、(a'): 「接続詞-後件. 前件」は不可能である。
- 接続助詞が用いられた場合、(A): 「前件-接続助詞. 後件」、(B): 「後件. 前件-接続助詞」という構造が考えられる²。複文における主節・従属節の倒置現象から、これらはどちらも発話対として認識可能である。

以上の考察は次の原則にまとめることができる[西澤 96].

原則 2 因果関係の表現に接続詞を用いた場合、因果関係の前件(原因/理由)および後件(結果)を述べている発話の位置関係は、次の条件を満たす。

1. 接続詞を含む発話は、その前方にある発話と組になって因果関係を表現する。前件/後件がどちらの発話となるかは接続詞の種類による。

² 「前件-接続助詞-後件」という談話構造も考えられるが、これは通常の複文であり、本稿の対象外となる。

2. 接続詞を含む発話とその後方にある発話は、これらの発話が意味内容として「前件(原因/理由)-後件(結果)」の順に並んでいる時に限り、組となって因果関係を表現することが可能となる場合がある。

一方、因果関係の表現に用いる接続助詞は、前件および後件を述べている発話の位置関係を特に制限しない。

以上の原則 1, 原則 2 をふまえて、実際の会話中に現れる発話対の談話構造を図 1 に例を示したコーパスを用いて調べてみる。コーパス中では、因果関係を表す接続詞として「だから/だって」、接続助詞(相当)として「から/のだから/のだもの」が用いられており、これらを含む発話が 1980 発話中 113 例存在した。そのうち本稿の対象は 97 例であり³、これらを次のように分類した。

- 原則 1 に基づいた分類。

まず、「隣接発話/そうでない発話」、「同一話者/異なる話者」という観点から、次のような分類がなされる⁴。

隣接する発話対	{	同一話者による発話対 … (A)
		異なる話者による発話対 … (*)
隣接しない発話対	{	同一話者による発話対 … (E)
		異なる話者による発話対 … (F)

さらに、(*)に分類される発話例はその特徴から次のように分類される。

(*)…	{	質問とその応答	因果関係を表さない … (B)
			因果関係を表す … (C)
		質問応答ではない(因果関係を表す) … (D)	

³ 次のような例は対象外とした。

1. 「頼むから、もう撮るの止めて下さい」など、通常の複文が発話されている場合。
2. 書きおこしの際に聞き取れていない部分が多く、発話自体の意味が不明な場合。

⁴ なお、ここでは、「発話の末尾が上がり調子“／”で終るかもしくは“／”を発話中に含む」または「『何で/どうして』などいわゆる疑問詞を含む」ような発話を質問文とした。

さらに、上記の分類に当てはまらないと判断される発話を (G) として分類することとする。

● 原則 2 に基づいた分類 [西澤 96].

- (a1) 前件. だから — 後件.
- (a2) 後件. だって — 前件.
- (a3) だって — 前件. 後件.
- (b1) 前件 — 接続助詞. 後件.
- (b2) 後件. 前件 — 接続助詞.

本稿で分類の対象とする 97 例のうち、上記の分類の (G) には 21 例が当てはまる。そこで、これらを除いた 76 例について、上記に従った分類を行なった結果が表 1 である。

表 1: コーパス中の発話対の分類

	(a1)	(a2)	(a3)	(b1)	(b2)	計
(A)	2	6	1	4	20	33
(B)	4	0	0	0	0	4
(C)	0	7	0	0	4	11
(D)	1	4	0	0	7	12
(E)	3	3	0	1	6	13
(F)	2	0	0	0	1	3
計	12	20	1	5	38	76

この結果から、次のようなことがいえる。まず、原則 1 に基づく分類からであるが、表 1 では、(A) が 33/76 で最も多い。これは、「隣接し、かつ同一話者である」という、聞き手側の理解度を強める二つの要素が重なった結果だと考えられる。逆に、(F) は「隣接せず、かつ話者が異なる」という、聞き手側の理解度が弱まる二つの要素が重なる。そのため、3/76 という少数の例しか存在しないものと考えられる。

次に、原則 2 に基づく分類からであるが、最も例が多いのは (b2) の 38/76 であり、続いて (a2) の 20/76 となっている。因果関係を表す複文では、後件は主節によって述べられ、こちらに重点を置くため主節と従属節を倒置する、という現象が見られるが、

これが (b2) に相当するものと考えられる。また、接続助詞を用いない場合でも、(a2) のように後件を先に述べる、という傾向があるものとも考えられる。また、原則 2 で例外的な場合とした (a3) は 1/76 しか存在せず、非常に特殊な例と考えられる。

ところで、(B) に属する例 (4/76) は因果関係を表さない例である。一例を次に示す。

- (1) a. F → A 肝試しとか騒いじゃいけないって意味じゃないんですか ↗
- b. A → F そうそうそう ↘
- c. M → A はい ↗
- d. A → M だから、肝試しとかうるさくするなってことだと思う ↘

この時、(1d) は前文に対する返答であるが、次の (2c) とは明らかに異なるニュアンスである。

- (2) a. A → B 何か、3種類くらい売っててこれが一番安かったみたい ↘
- b. F → B 人体実験やんないの ↗ 今日 ↘
- c. B → F だってもう眠いもん ↘
- d. B → F 何か最近酒弱いみたい ↘
- e. A → B 俺も弱くなった ↘

(2c) では前文の質問に対して、「(やらない、) だってもう眠いもん」のように後件を省略した形で返答している。この後件は前文の意味内容と関係があるため、(2b) - (2c) は (間接的に) 因果関係を表す発話対であると考えられる。一方、(1d) は前文の質問に対して「それは前に説明したはずだが、覚えていないのか」というような態度を表明していると考えられる。このような接続助詞の用法は、コーパス中では「だから」の場合のみであり、質問に対して「だから」で応答すると必ずこの用法になる⁵ことから、「だから」の用法について次のことがいえるのではないかと考えられる。

仮定 1 質問文に対して、「だから」を用いて返答した場合、質問文と返答文の意味内容は因果関係を表さない。

⁵表 1 で (a1)-(B) が 4 例であるのに対し (a1)-(C) が 0 例であることが、このことを示している。

3 発話対の推定

3.1 推定の手順

前節で検討した結果を元に、ここでは、因果関係を表す接続(助)詞によってできる発話対を推定するための方法について検討する。これについては、すでに[西澤96]などで同一人物による発話対⁶を対象とした方法が述べられているが、今回はそれを参考にし、異なる話者による発話対をも対象に含んだ手順についての検討を行なう。

基本的な手順としては、前節で述べた原則や表1の結果をもとに、最も発話対を形成しやすい場合からコーパス内をチェックしていく、ということになる。ここでは、次のような発話対推定の手順を仮定する。

仮定2 因果関係を表すとされる接続(助)詞による発話対をコーパス中で発見(推定)するための手順を次のように仮定する。

1. 該当する接続(助)詞⁷が含まれる発話をコーパス中から発見する。この発話を U_1 (話者 S_1 , 聞き手 H_1) とし、直前の発話を U_0 (話者 S_0 , 聞き手 H_0)、直後の発話を U_2 (話者 S_2 , 聞き手 H_2) とする。
2. 接続(助)詞が「だから/だって」の場合でかつ $S_0 = S_1$ ならば、 U_0 と U_1 は発話対となる。また、接続(助)詞が「から/のだから/のだもの」の場合でかつ $S_0 = S_1$ ならば U_0 と U_1 が、 $S_1 = S_2$ ならば U_1 と U_2 が発話対となる。
3. 2.にあてはまらない場合: 接続(助)詞が「だから」以外の場合でかつ U_0 が質問文ならば、 U_0 と U_1 は発話対となる⁸。一方、接続助詞が「だから」の場合でかつ U_0 が質問文の場合も、 U_0 と U_1 は発話対となるが、この場合の発話対は因果関係を表さない。

⁶表1の(A)および(E)にあたる。

⁷本稿では、「だから/だって/から/のだから/のだもの」を想定する。

⁸なお、この場合は、発話対が因果関係を直接述べるのではなく、間接的に述べるものがほとんどである。

ただし、両者ともに、 U_0 があいづちや単なる問い返しのような場合⁹は例外となる。

4. 3.にあてはまらない場合: 発話 U_0 において、 $H_0 = S_1$, $H_0 = H_1$, $S_0 = H_1$ のいずれかであるならば、 U_0 と U_1 は発話対となる。ただし、この場合も、 U_0 があいづちや単なる問い返しのような場合は例外となる。
5. 4.にあてはまらない場合: 発話 U_1 から最も近い位置にある、話者 S_1 による発話をコーパスから探し、それを発話対の対象とする。この時、接続(助)詞が「だから/だって」の場合は、発話 U_1 から前方へのみの検索を行ない、「から/のだから/のだもの」の場合には前方および後方への検索を行なう。

この手順では、まず2.の手順によって、発話同士が隣接しかつ同一人物による発話であるような発話対が発見される。次に、3.の手順によって、発話同士が隣接し、かつそれらが質問とそれへの返答になっているような発話対が発見される。次の手順4.では、発話対の話者および聞き手がそれぞれの発話同士で関係している(交替もしくは聞き手が同一、など)場合が対象となる。最後に、5.の手順で、隣接していない同一話者による発話対が発見される。また、原則2を考慮し、接続(助)詞が「だから/だって」の場合には前方へのみ、「から/のだから/のだもの」の場合には前方および後方への検索を行なうこととしている¹⁰。

3.2 推定手順についての考察

上のように仮定した手順を用いて、実際にコーパスを手作業により検索した結果、次のような結果となった。

- 本稿で検討対象としている発話97例のうち、表1の(A)～(F)に分類される76例がこの手順の適用対象であり、適用可能率は76/97つまり78.4%である。

⁹「ああ\」、「うん\」、「何\」のような発話を指すものとする。

¹⁰原則2では「だって」の場合の例外について述べているが、コーパス中にも1例しか存在せず、非常に特殊な例であると考えられるため、仮定2の手順では考慮していない。

- 76 例に対してこの手順を適用した結果、人間による判断と同一の発話対が推定できた例は 59 例であった。さらに残り 17 例のうち、後述する、[西澤 96]にあるような補足手順を用いることによって、11 例が推定可能となった。これにより、適用成功率は 70/76 つまり 92.1% である。

人間によって発話対の認識が可能な発話を対象とした場合、仮定 2 の手順は、59/76 つまり 77.6% の割合で正しい (人間の判断と同一の) 発話対を推定する。さらにいくつかの補足手順を併用することにより、最終的には 92.1% の割合で正しい発話対が認識される。この、補足手順が必要な例とは、次のような場合である。

まず、(3) のように、接続助詞を含む発話の前後に、同一話者による発話が存在する場合である。

- (3) a. C → * いや、笑わないでよ \
- b. C → * これ、真剣なんだから \
- c. C → * 別に笑ってもいいんだけど \

これは、仮定 2 の手順 2 が適用されるべき場合であるが、発話対が (3b) とその直前の発話なのかそれとも直後の発話なのかの判断は仮定 2 では行なえない。しかし、これらの例では、(3a) のように「～よ」などの終助詞によって話者が主張の態度を明確にしておき、それに対する理由を続けて「～のだから」などと発話していると考えられる。そこで、次のような補足手順を仮定する。

接続助詞を含む発話の直前と直後に、同一話者による発話隣接している場合、「～よ / よね」などの表現によって、主張を述べていることが明確になっている発話を発話対の対象とする。

この補足手順により、2 例が推定可能となる。また、(3) と類似した次のような場合も存在した。

- (4) a. C → C 張ってなかったかね →
- b. C → C あっいや、ベイ単騎は絶対ないわけだ \
- c. C → C もう 3 枚出てるから \
- d. C → C 張ってなかった、張ってなかった \

これも (3) の場合と同様に、接続助詞を含む発話と発話対をなすのは直前の発話なのか直後の発話なのかの判断が仮定 2 では行なえない。この場合は、次のような補足手順が有効であると考えられる。

接続助詞を含む発話の直前と直後に、同一話者による発話隣接している場合、「～わけだ」など因果関係を明示するような語句が用いられている発話を発話対の対象とする。

この補足手順に該当する発話例は、上で述べた 1 例存在した。さらに、[西澤 96]で述べられている補足手順のうち、いくつかの手順については有効であることが確認された。これら補足手順を以下に述べる。

検索の途中で、「それで / そして」など話題を先へ進める力を持つ語を含む発話が発見された場合、その発話から先への検索は行なわなくてもよい。

話し手 S_1 と同一人物による発話を検索の途中で、 S_1 による、「から / のだから / のだもの」を含む発話が発見された場合は、その発話を発話対の対象とはせず、さらに検索を進める。

発話 U_1 (話者 S_1 , 聞き手 H_1) の直前と直後に、 U_1 による発話が存在する場合は、それのうち聞き手が H_1 である発話が発話対の対象となる。

以上の補足手順の適用により、適用成功率は最終的に 92.1% となる。仮定 2 の手順およびいくつかの補足手順は、発話隣接しているか、また話者や聞き手が同一であるかどうか、など、発話の表層情報を利用するものであることから、このような発話対認識を計算機上で行なう場合、本稿で述べた手順は非常に有効な方法となるのではないかと考えられる。

4 おわりに

本稿では、因果関係を表すとされる接続 (助) 詞によって表される発話対を対象とし、その談話構造について原則 1, 原則 2 を述べた。さらにそれをもとにし

て、発話対となる部分を会話コーパス全体から推定する手順について、仮定 2 で述べた手順およびいくつかの補足手順を示し、これが有効に機能することを確認した。

ところで、仮定 2 の手順では、人間によって認識される発話対を推定しきれない場合が、以下のように 6/76 つまり 7.9% 存在する。

- コーパスへの書きおこしの際に、話し手や聞き手の情報が得られなかったために、情報不足となっている例 … 2 例。
- 原則 2 で例外とした、「だって—前件. 後件。」というタイプの発話対(表 1 の(a3)にあたる) … 1 例。
- 発話の意味解析が必要と思われる例 … 3 例。

これは例えば次のような発話例である。

- (5) a. E → B でもさー、テニスとかいつも二日酔いで寝てない / 午前中 \
- b. B → E 俺は寝てないんだけど \
- c. F → * だから 1面なんです \

この例では、第一文と第三文とが発話対となると考えられるが、話し手や聞き手などの共通性がこれらの間にみられないことや、第二文が挿入されていることなどから、仮定 2 やこれを補足する手順のみでは扱えない。

また、今回は、人間によっても発話対を認識できない場合をあらかじめ排除した上で仮定 2 などの手順を適用しているが、排除自体は手作業で行なっている。これに関する手順の検討なども、実際に発話対認識を計算機上で行なう際には必要となると考えられ、この点については今後の課題である。

謝辞

本研究は、文部省科学研究費重点領域研究「音声言語」の補助を受けていることを記し、御協力いただいた関係各位に感謝致します。

参考文献

- [Hob90] Jerry R. Hobbs. *LITERATURE AND COGNITION*, pp. 83-114. No. 21 in CSLI Lecture Notes. CSLI, 1990.
- [高野 93] 高野敦子, 柏岡秀紀, 平井誠, 北橋忠宏. 発話対の認識手法について — 質問とその応答. 情報処理学会研究報告 93-NL-94, pp. 17-24. 情報処理学会, 1993.
- [田中 92] 田中智博, 林良彦. 文の接続パターンに基づく日本語テキスト構造の解析. 情報処理学会研究報告 92-NL-89, pp. 1-8. 情報処理学会, 1992.
- [西澤 96] 西澤信一郎, 中川裕志. 日本語の自由会話における発話間の因果関係の推定. 言語処理学会 第 2 回年次大会 発表論文集, pp. 341-344. 言語処理学会, 1996.
- [福本 92] 福本淳一, 安原宏. 文の接続関係解析に基づく文章構造解析. 情報処理学会研究報告 92-NL-88, pp. 9-16. 情報処理学会, 1992.
- [三上 72] 三上章. 現代語法序説. くろしお出版, 1972. 刀江書院発行 (1953) を復刊.
- [森田 80] 森田良行. 基礎日本語 2. 角川小辞典. 角川書店, 1980.