

電子メールを利用したコミュニケーションにおける 討議の流れの自動抽出法について

村越 広亨† 島津 明† 落水 浩一郎†

†北陸先端科学技術大学院大学 情報科学研究科
〒 923-1292 石川県能美郡辰口町旭台 1-1

Email: {murakosi, shimazu, ochimizu}@jaist.ac.jp

あらまし

本研究では、ソフトウェア開発などの共同作業で利用される電子メールによるコミュニケーションに着目し、討議の流れの参照を容易にするための構造モデルの構築、自動抽出法の提案を行なう。具体的には、コントリビューションツリーの考え方を利用し、メール群を入力として、発話ペアを話題毎に構築することにより発話を整理、保存し、討議の流れを表現した構造木モデルを提案する。仕様書レビュー作業の事例の構造木を構築し、構造木を構成する発話の言語的特徴を分析する。その結果から、討議の構造木の自動抽出法の検討、提案を行なう。

The Way of Automatic Extraction of Deliberation Stream from Email Messages

Hiroyuki MURAKOSHI† Akira SHIMAZU† Koichiro OCHIMIZU†

†School of Information Science,
Japan Advanced Institute of Science and Technology,
1-1, Asahidai, Nomi-gun, Ishikawa, 923-1292 Japan

Email: {murakosi, shimazu, ochimizu}@jaist.ac.jp

Abstract

In this paper, we focus on communication using email in a cooperative work such as software development. We propose a model of deliberation stream making it easy for a speaker and listeners to recognize utterances related to the preceding ones. Then we propose the way of automatic extraction of the deliberation stream from email messages. In this model, the concept of *contribution tree* is adopted, and adjacency pairs of utterances are composed every topic. They are connected so that structure trees showing deliberation stream are constructed. We construct the structure trees of four examples of software specification reviews and analyze the linguistic feature of utterances. From the results of our analysis, we examine and propose the way of automatic extraction of the structure tree of deliberation.

1 はじめに

電子メールを利用したコミュニケーションでは、対面の場合と同様な会話のターンテキングを利用できず、聞き手が話し手の発話を理解していることを伝達しづらいと思われる。また、聞き手が自分の発話を理解しているかどうかを認識するための手がかりも少ないといえる。そのような状況では、対話者間の会話に関する共通理解が阻害され、認識の不一致の発生が予想される [1]。

共同ソフトウェア開発において、コミュニケーションにおける作業員間の認識の不一致は、ソフトウェア最終成果物に悪影響を及ぼすことが考えられる。過去の発話を話題毎に分類、整理し、発話意図の聞き手の理解状況を確認でき、適切な返答を行なっているかどうかといった討議の流れを参照することができれば、話題の終結に関する作業員間の認識の不一致を回避したり、同じ話題の議論の繰り返しが少なくなるといった効果があると思われる [2]。そこで、コントリビューションツリー [3] の考え方を利用し、メール群を入力として、発話ペアを話題毎に構築することにより発話を整理、保存し、討議の流れを表現した討議の構造木モデルを提案する。さらに、ソフトウェア仕様書レビュー作業の事例から、討議の構造木を構築し、構造木を構成する発話の言語的な特徴を分析する。その結果から、構造木の自動抽出法の検討、提案を行なう。

2 電子メールによる討議の構造

2.1 討議の流れ

電子メールを利用したコミュニケーションは、Text型で、かつ対面の場合と同様なターンテキングを利用できないため、1対1の発話のやりとりが行なわれないことがある。そのような状況では、複数の発話が相手から一度に投げかけられ、並列に話題が進められることが考えられる。またメールには、話題毎にサブジェクトがつけられることが多く、1メール中には、複数の発話が存在することが多いと思われる。また、同じサブジェクトのメールが連続して行なわれるとは限らず、作業員が交互にメールのやりとりを行なうとも限らない。図1に、電子メールによる討議の流れを示し、発話、トピック、サブトピックを以下に定義する。

トピック：メールのサブジェクト毎に議論される内容のこと。

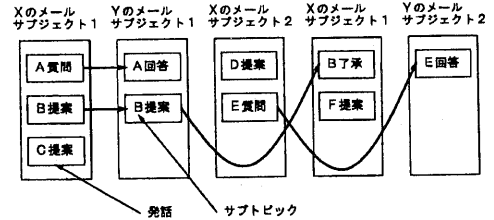


図1: 電子メールによる討議の流れ

サブトピック：メール中に複数存在し、トピックを細分化したもの。

発話：同一メール内で、同じサブトピックを言及している文章のまとまりのこと。自分の意見や提案などを示している。図1では、横に長い長方形で表現される。

例えば、1では、サブトピックAの質問(A質問)とサブトピックAの回答(A回答)といったような発話のペアが構成されている。発話のペアは、発話の隣接ペア [4] から決定される。質問、回答といった行為による発話の意味づけは、文献 [5] を参考にしている。

2.2 連続した発話の分類

図1に示すように、電子メールを利用したコミュニケーションでは、Text型で、かつ対面の場合と同様なターンテキングを利用できないため、1メール中に複数の発話の存在が予想される。複数の発話には、相手への返答や新たなサブトピックに関する意見や提案などが含まれる。また、複数の発話間に関連のある状況や、返答の前に、返答に関連する発話が行なわれる状況が考えられる。連続した発話を以下に定義する。

連続した発話：メール中に含まれる複数の発話のうちで、隣接した発話同士のこと。

メール中の文書が、意味的つながりをもったテキスト構造ならば、連続した発話は重要であり、発話間の関連と密接な関係があると思われる [6]。同一メールにおける連続した発話の前者の発話と後者の発話は、それぞれ表1のようなパターン(P)に分類することができる。

3 討議の流れを表現する構造木モデル

3.1 コントリビューションツリー

図1のような電子メールを利用したコミュニケーションの流れを保存、整理するためのモデルとして、

表 1: 同一メール中の連続した発話のパターン

パターンの種類	前者の発話	後者の発話
P1 (連続非返答)	非返答	非返答, 前者の発話と非関連
P2 (連続非返答)	非返答	非返答, 前者の発話と関連
P3 (非返答-返答)	非返答	返答, 前者の発話と非関連
P4 (非返答-返答)	非返答	返答, 前者の発話と関連
P5 (返答-非返答)	返答	非返答, 前者の発話と非関連
P6 (返答-非返答)	返答	非返答, 前者の発話と関連
P7 (連続返答)	返答	返答, 前者の発話と非関連
P8 (連続返答)	返答	返答, 前者の発話と関連

コントリビューションツリー [3] の考え方を利用する。コントリビューション (Contribution) とは、話し手と聞き手による会話のユニットのことである。会話参加者がコントリビューションを形成するには、話し手の発話の意図を聞き手が理解しているかどうかといった事実 (会話参加者間の共通認識) を話し手が獲得する必要がある。コントリビューション (C) は、Presentation Phase (Pr) と Acceptance Phase (Ac) から形成される。

Pr : 作業 X が作業 Y に、自分の意志を伝達するための発話 u を行なうフェーズのこと。

Ac : 自分が発話 u を理解したかどうかを、作業 X が理解できると思われる行動 e を作業 Y が作業 X に与えるフェーズのこと。

コントリビューションを形成する 2 つのフェーズのペアは、発話の隣接ペア [4] から決定される。例えば、 C を構成する Pr (先行発話) が質問タイプの発話の場合、 Ac (後続発話) は回答タイプの発話となる。

コントリビューションツリー (Contribution Tree) とは、会話参加者によるコントリビューション (C) がツリー状に構築されたものである。図 2 に基本的なコントリビューションツリーの例を示す。このツリーには、以下の特徴がある。

- すべての発話は、ツリーを構成するコントリビューションに属するプレゼンテーションフェーズに対応づけることができる。
- ほとんどのコントリビューションは、同一レベルのコントリビューションの開始によって完結する。
- 発話 A の内容が作業 Y に理解されている場合、図 2 で示されるように、発話 B が Ac であるとともに、次のコントリビューションの Pr となる。

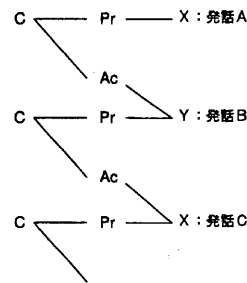


図 2: コントリビューションツリー

3.2 討議の構造木モデルの提案

コントリビューションツリーの考え方を利用した討議の構造木モデルを提案する。この構造木モデルでは、同じトピックの議論をしている同一サブジェクト毎に発話を整理することにする。

3.2.1 サフィックスの導入

C , Pr , Ac にサフィックスを導入する¹。サフィックスの第 1 ラベルをサブジェクト番号とし、それを以下に定義する。

サブジェクト番号: サブジェクト毎に割り当てた番号のことで、導入された順番 (1, 2, 3, ...) でナンバリングされる。

図 3 に構造木を構成する C のサフィックスを示す。 $C_{x,y,z,\dots}$: x は、サブジェクト番号を示す。 y , z とサフィックスが深くなると、コントリビューションのレベルが深くなることを示す。 y 以降のサフィックスの値は、コントリビューションが導入された順番を示す。

C を構成する Pr と Ac のサフィックスは、 C と同様のものとする。

¹ 文献 [3] では、コントリビューション、プレゼンテーションフェーズ、アクセプタンフェーズのサフィックスが定義されていない。

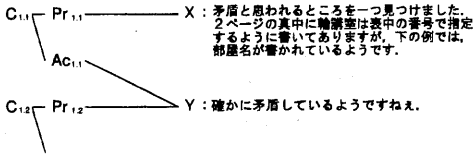
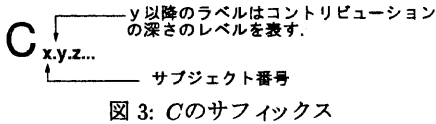


図 4: 相手の発話の理解を前提とした返答の場合の発話ペアの例

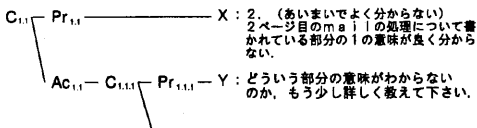


図 5: 相手の発話に関する質問や疑問の場合の発話ペアの例

3.2.2 発話ペアの構築

作業者同士がお互いの存在を認識し、会話を行なっている状況では、新たな話題を相手に投げかけたり、相手の発話に対して返答を行なっている。このような発話ペアの構築は会話の基本である。相手の発話に対する返答は、以下のような場合が考えられる。

1. 相手の発話の理解を前提とした返答を行なう場合。
2. 相手の発話を十分に理解できず、相手の発話に関する質問や疑問を示す返答を行なう場合。

図 4に上記 1 の場合、図 5に上記 2 の場合の構造木の例を示す。同じ話題毎に発話ペアの構造木を組合わせていき、討議全体の構造木が構築される。

図 4は、作業者 X がサブジェクト番号 1 の最初の発話を行ない、作業者 Y が、その発話の理解を前提とした発話を行なっている場合の発話ペアの構造木を示す。つまり、作業者 X によるプレゼンテーションフェーズの発話の内容を作業者 Y が十分に理解し、アクセプタンスフェーズの発話は、作業者 X による発話が理解されたことを前提にした発話となっている。このような場合の完結したコントリビューションを構成する Ac は、次の同一レベルのコントリビューションの Pr となる。

図 5は、作業者 X がサブジェクト番号 1 の最初の発話を行ない、作業者 Y が、その発話の内容に関する質問を行なっている場合の発話ペアの構造木を

表 2: 対を構成する語

先行発話(問いかけ)	後続発話(応答)
思いませんか?	そう思います
どうでしょうか	良いと思います
変ですね	確かに変ですね
終わりにしますか	終わりにしましょう

示す。つまり、作業者 X によるプレゼンテーションフェーズの発話内容を作業者 Y が十分に理解していない場合、アクセプタンスフェーズの発話 B は、発話内容に関する質問、疑問を示すような発話となっている。

4 事例による討議構造木の構築と分析

共同作業における電子メールを利用した 4 つの討議事例から、3.2 節でモデル化した討議構造木を構築し、その構造木の特徴を示す。事例は 2 人の作業者による電子メールを利用したソフトウェア仕様のレビュー作業で、平均で 31.5 のメールのやりとりがあった。

4.1 発話の特徴

メール中の文書を発話に分類するための手がかりを得るために、発話ペアを構成する発話の特徴を分析したところ、空行で分けられた文章のまとまりが発話単位であることが多かった。また発話の最低単位は一文であり、一文中に空行があっても一文が発話であった。

4.2 発話ペアの特徴

事例による構造木の発話ペアを分析したところ、2 つのフェーズ (Pr と Ac) による発話ペアは、引用文から推測できる発話 (Pr) と、引用文直後の発話 (Ac) との間で構築されることが多かった。以下に、発話ペアを構成する 2 つのフェーズの発話を分析する。

4.2.1 対を構成する語

発話ペアを構成する先行発話と後続発話の先頭文中から、先行発話中の相手への問いかけと、後続発話の先頭文中に、それに対応した同意を表すような語のペアがあることが分かった。それは、発話ペアを見つけるための手がかりになるとと思われる、対を構成する語の例を表 2 に示す。対を構成することを示すような手がかり語により、後続発話が、先行発

表 3: 返答要求を示す語の例

語尾	例
か	思いませんか
かな	私の提案で良いかな
ましょう	移りましょう
でしょう	どうでしょう
よ	時間がずれているよ
ね	決める必要がありますね

話の理解を前提として行なわれ、適切な発話であるかどうかを識別することが可能になると考えられる。

4.2.2 プレゼンテーションフェーズの発話の特徴

プレゼンテーションフェーズでは、新たなサブトピックの議論を開始する発話と、相手への返答からサブトピックを継続する発話に分類することができる。プレゼンテーションフェーズの発話の特徴として、表3に示すように、発話中の文の語尾に意見や疑問、提案を示す語が含まれることが多く、特に、相手への返答ではない場合に顕著であった。

4.2.3 アクセプトフェーズの発話の特徴

アクセプトフェーズの発話は、以下の3種類に分類することができた。

1. サブトピックを終結する発話 (相手の発話の理解を前提とした返答)。
2. サブトピックを終結させない発話。
 - (a) 相手の発話の理解を前提とした返答。
 - (b) 相手の発話を理解を前提としていない返答。

表4に示すように、サブトピックを終結し、相手の発話の理解を前提とした発話には、発話の先頭文中に、相手の発話の理解を前提として同意を示したり、相手の発話に対する感嘆を表現する語が含まれていた。また、発話の先頭文以外の文の語尾に、表3で示すような意見や疑問、提案を示す語が含まれることがなかった。サブトピックを終結させず、相手の発話の理解を前提としていない発話は、相手の発話の内容についての質問があった。

4.3 連続した発話の構造木の特徴

同一メール中に複数の発話が含まれる場合、4.2節における発話ペア以外の構造が考えられる。表1の連続した発話の各パターン2の構造木を分析し、それぞれの構造木の言語的な特徴を示す。ただし、4つの事例からは、後者の発話が前者の発話と関連があ

表 4: 適切な返答を示す語の例

同意	感嘆
賛成です	なるほど
決定です	するどいですね
そのとおりですね	

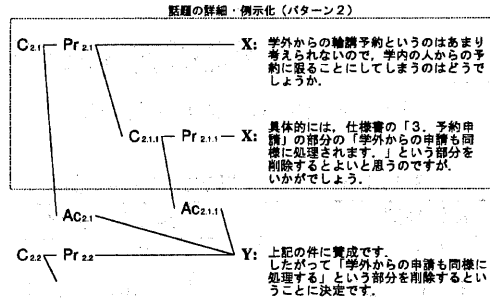


図 6: 話題の詳細・例示化の例

るような連続した返答 (パターン8) を見つけることができなかった。

4.3.1 前後の関連性を捉える特徴

表1で分類した連続した発話のパターン2, 6では、非返答である後者の発話が前者の発話と関連しており、パターン1.5では、後者の発話に関連していない。前後が関連した場合の構造木は、図6や図7のような構造となった。また、前後が関連していない構造木は図8のような構造となった。以下に、それぞれの構造木の例と言語的特徴を示す。

話題の詳細・例示化 図6は、後者の発話内容が、前者の発話内容の詳細化や例示化を示しているような連続した発話の構造木を示す。図6では、作業員Xによって、サブジェクト番号2の連続した発話 (パターン2) が行なわれ、後者の発話が前者の発話を具体的に説明する発話となっている。ただし、図6の場合、後者の発話は、前者の発話と同一レベルのコントリビューションの発話とはならない。

このような構造木を構成する発話を分析したところ、後者の発話の文頭に、「具体的には」といったような、詳細化や例示化を示すような接続詞がみられたり、後者の発話の先頭文に、「これ」といったような指示代名詞が含まれる傾向にあった。

前提をおいた話題展開 図7は、前者の発話、後者の発話の前提を意味しているような連続した発話の構造木を示す。図7では、サブジェクト番号2の

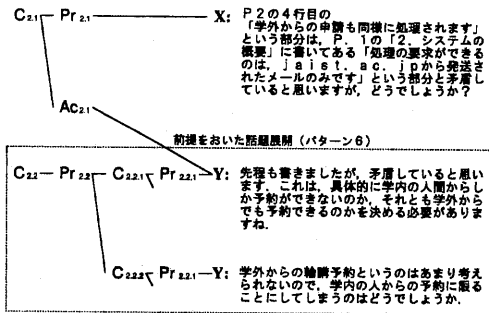


図 7: 前提をおいた話題展開の例

表 5: 前提をおいた話題展開を示す語

接続詞例	指示代名詞例
それで, で, だから (順接)	この,
さらに (追加)	これは, これに

新たなサブピックの発話を作業員 X が行った後、作業員 Y が、2つの連続した発話 (パターン 6) を行っている場合の構造木を示す。連続した発話の前者の発話は、作業員 X による発話の返答になっている。この返答は、相手の発話の理解を前提とし、サブピックを終結させない発話である。後者の発話は、前者の発話を前提として行なわれている。よって、前者の発話と後者の発話をまとめて、作業員 X による発話と同レベルのコントリビューションを構成している。

この構造木を構成する発話を分析したところ、表 5 に示すように、後者の発話の文頭に接続詞がみられたり、後者の発話の先頭文に指示代名詞がみられる傾向にあった。

話題の並列化 図 8 は、連続した発話の前者の発話と後者の発話の間に、詳細化や前提といった関連がなかったり、前者の発話がサブピックを終結する返答である場合の構造木を示す。図 8 では、作業員 X がサブジェクト番号 3 の新たなサブピックの発話を行った後、作業員 Y が連続した発話 (パターン 5) を行ない、作業員 X による発話に対する返答を行なっている場合の構造木を示す。図 8 の場合、作業員 Y による連続した発話の前者の発話は、サブピックを終結する返答で、後者の発話との間には関連がない。また、前者の発話がサブピックを終結する返答である場合、後者の発話は関連した発話ではないので、連続した発話の構造木は図 8 のような構造となる。

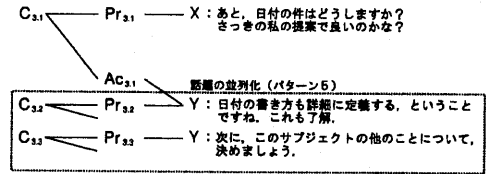


図 8: 話題の並列化の例

表 6: 新たなサブピックを示す語

	接続詞例
話題の追加	それから, あと, それと
話題の転換	次に, 次は

さらに、同一メール中に、前者の発話の直前の発話 (以後、前々者の発話と呼ぶ) が存在し、その発話が前者の発話と関連がある場合、以下の 2 種類の構造木が存在していた。

1. 後者の発話が前々者の発話を前提としたような発話の場合、前者の発話と後者の発話による並列化の構造木。
2. 後者の発話が前々者の発話を前提としたような発話でない場合、前々者の発話と後者の発話による並列化の構造木。

話題の並列化の構造木を構成する発話を分析したところ、サブピックの議論を並列に進めることを明示的に指示するようなメタ的な発話が存在していた。これはパターン 1 の場合に適用できる。メタ的発話には、サブピックの並列展開を示すような、「列挙します」「挙げます」といった語が存在していた。さらに、表 6 に示すように、新たなサブピックの展開を表す接続詞が発話の先頭文にみられた (パターン 5 の場合)。

4.3.2 返答を捉える特徴

表 1 で分類した連続した発話のパターン 4 では、後者の返答が非返答である前者の発話と関連しており、パターン 3 では、前者の発話と関連していない。またパターン 7 では、後者の返答が前者の返答と関連していない。関連した発話を前提とした場合の構造木は、図 9 のような構造となった。また、関連した発話の前提がない場合の構造木は図 10 となった。連続した返答の場合の構造木は図 11 のような構造となった。以下に、それぞれの構造木の例と言語的特徴を示す。

前提返答 図 9 は、相手の発話に対して連続した発話が行なわれるときに、最初に相手の返答に対する

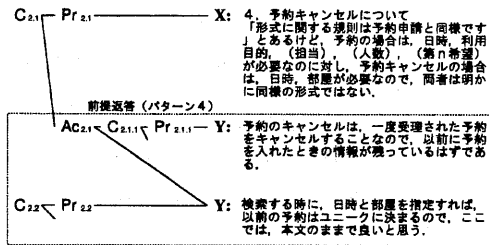


図 9: 前提返答の例

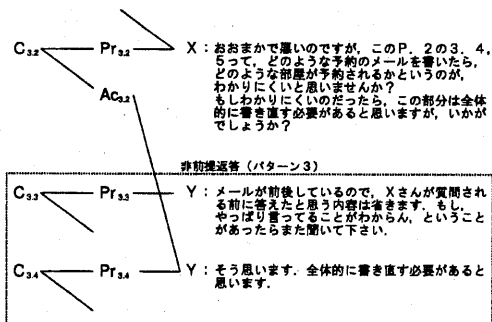


図 10: 非前提返答の例

回答を行なうのではなく、理由を述べてから、相手の返答に対する回答を行なう場合の構造木を示す。図9では、作業者Xがサブジェクト番号2の新たなサブピックの発話を行なった後、作業者Yが連続した発話(パターン4)を行ない、作業者Xによる発話の返答を行なっている場合の構造木を示す。図9では、連続した発話の前者の発話が、後者の発話の理由となっている。このパターンは頻度が少なく、言語的特徴を見つけることができなかった。

非前提返答 図10は、連続した発話が行なわれるときに、後者の発話が返答の発話で、前者の発話が後者の発話と関連がない場合の構造木を示す。図10では、作業者Xがサブジェクト番号3の発話を行なった後、作業者Yが連続した発話(パターン3)を行ない、前者の発話が、作業手順などを示すようなメタ的な発話で、後者の発話が作業者Xの発話の返答である場合の構造木を示す。この構造木を構成する連続した発話同士の特徴として、関連を示すような語が存在していなかった。

連続返答 図11は、複数の返答が行なわれるときに、それぞれの返答が発話ペアを構築する場合の構造木を示す。図11では、作業者Xがサブジェクト番号3の新たなサブピックの発話を連続して行なった後、作業者Yがそれぞれ発話に対応した返答を行なって

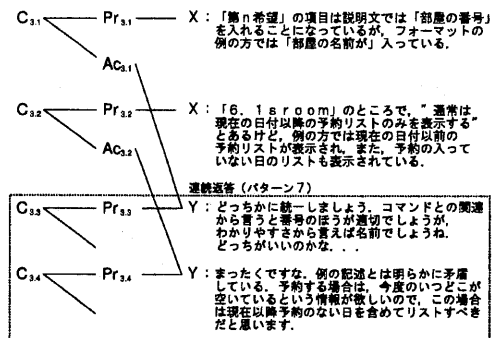


図 11: パターン6の例

いる場合(パターン7)の構造木を示す。この構造木では、それぞれの返答ごとに、適切なプレゼンテーションフェーズの発話を抽出し、発話ペアの構造木を構築することが必要となる。

5 討議構造木の自動抽出法

事例から構築した討議の構造木を構成する発話の分析を利用し、構造木の自動抽出法の検討、提案を行なう。

5.1 文書からの発話抽出

空行で分けられた文章のまとまりを発話単位とし、メール中の文書から発話を抽出する。ただし、発話の最低単位を一文とし、一文中の空行は無視することにする。

5.2 構造木抽出アルゴリズム

メール毎に文書から抽出した発話群を入力とし、討議の構造木を抽出するための方法を提案する。図12と図13は、同一メール中の発話の構造木を構築するためのフローチャートを表している。図12は、同一メール中の発話の一つの場合の構造木の構築フローを示している。図13は、図12のフローの続きであり、同一メール中の発話が複数ある場合の構造木の構築フローを示している。図12、図13の処理2から処理8では、以下に示すように、4.2節や4.3節で分析した構造木を構成する発話の言語の手がかりなどを利用する。

発話ペア構築: 処理2, 6, 7, 8では、返答の適切なプレゼンテーションフェーズを決定し、発話ペアの構造木の構築を行なう。この場合、引用文などを手がかりとして、適切なプレゼンテーションを決定する。また、表2や表4のような手が

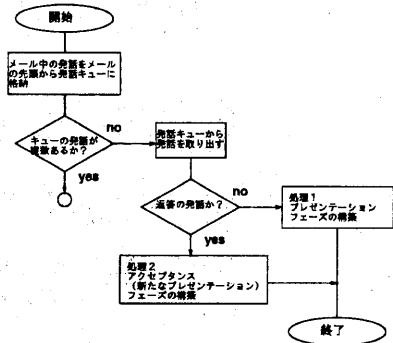


図 12: 同一メール中の発話が 1 つの場合

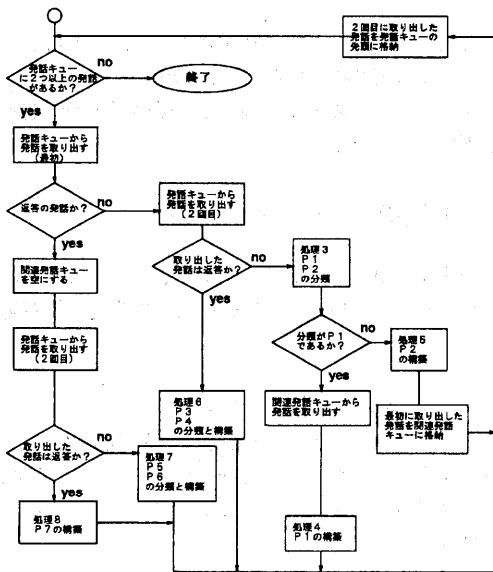


図 13: 同一メール中の発話が複数存在する場合

かり語により、相手の発話の理解を前提とした返答か否か決定し、図 4あるいは図 5の構造木を構築する。

処理 3: P1(パターン1)とP2の分類は、

- 連続した発話の後者の発話の先頭文に、表 5に示すような接続詞や、先頭文中に指示代名詞が含まれるかどうか?
- 前者の発話が、サブトピックを終結する返答であるかどうか?
- 後者の発話の先頭文に、話題の転換を示すような手がかり語があるかどうか?

といったことから決定する。

処理 4: 連続した発話の前者の発話の直前の発話(以後、前々者の発話と呼ぶ)が存在し、その発話が前者の発話と関連がある場合に、

- 後者の発話の前々者の発話を前提とした発話ならば、前者の発話と後者の発話による構造木の構築
- 後者の発話の前々者の発話を前提としたような発話でないならば、前々者の発話と後者の発話による構造木の構築を行なう。

処理 5: P2の構造木は、後者の発話の先頭文の接続詞の種類を手がかりとして決定する。

処理 6: P3とP4の分類は、後者の発話が引用文の直後にあるかどうかで決定する。直後にある場合はP4とする。

処理 7: P5とP6の分類は処理3と同様に行ない、P6の構築は処理5と同様に行なう。

6 結論と今後の課題

本研究では、コントリビューションツリーの考え方を利用し、電子メールによるコミュニケーションの討議の流れを表現する構造モデルを提案した。ソフトウェア仕様レビュー作業の事例から構造木を構築し、構造木を構成する発話の言語的特徴を分析することにより、構造木を自動抽出するための手法の検討、提案を行なった。今後の課題として、以下の点を行なう予定である。

- 構造木を構成する発話の言語的特徴のさらなる分析
- 提案した手法による構造木自動抽出プログラムの実装、評価

参考文献

- [1] 村越広享, 落水浩一郎. 電子メールを利用した共同作業における会話のコヒーレンスと話題の完結度の関係. *コンピュータソフトウェア*, Vol. 15, No. 3, pp. 50-53, 1998.
- [2] 桑名栄二. ソフトウェア履歴利用の研究動向. *電子情報通信学会誌*, Vol. 77, No. 5, pp. 531-538, May 1994.
- [3] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, Vol. 13, No. 2, pp. 259-294, 1989.
- [4] Charles Goodwin and John Heritage. Conversation analysis. *Annual Reviews Anthropol*, Vol. 19, pp. 283-307, 1990.
- [5] J.R. Searle. *Speech Acts*. Cambridge University Press, 1969.
- [6] M. A. K. Halliday and Ruqaiya Hasan. *Coherence in English*. Longman, 1976.