

## 旅行案内電話対話のモデル化における 基盤化ネットワークの拡張と適用

### An expansion and application of the grounding network for modeling telephone-based tour guidance dialogue

水上 悦雄<sup>†</sup>, 柏岡 秀紀<sup>†</sup>  
Etsuo Mizukami, Hideki Kashioka

<sup>†</sup> 情報通信研究機構  
National Institute of Information and Communications Technology  
etsuo.mizukami@nict.go.jp

#### Abstract

In goal-oriented or task-oriented conversations, the participants have to share many things to achieve their collaboration. The ideas of *common ground*, *shared knowledge* and the similar concepts are important to understand the process of the achievement. In this study, to model the guide-tourist dialogue, we proposed an idea of *extended grounding networks* by introducing a concept of *contribution topic*, and applied it to the data.

**Keywords** — Extended Grounding Network, Contribution topic, Dialogue Model

#### 1. はじめに

人の対話戦略、特に、旅行案内における熟練ガイド、会議や話し合いにおけるモデレータ、ファシリテータとよばれるプロの司会者、カウンセリング場面におけるプロのカウンセラーなどの対話技術をモデル化し、音声対話システムなどの計算機システム上で実現するために、様々なアプローチでの研究が進められてきた。そのような対話現場で求められる技術とは、利用者や参加者の、時に曖昧で不完全、順不同かつ予測不能な言動に対しても柔軟に対応し、最終的にそれぞれの対話のゴールに向かって対話を導くことであり、人ならではの対話技術とも言える。そのような複雑な対話戦略のモデル化に関して、近年では、大規模なタグ付コーパスなどを利用した、統計的なフレームワークを用いた対話制御手法が主流となってきている[1][2][3]。これらの統計的モデルの利点は、IF/THENルールで書き尽くせない複雑な対話状況も、統計的にもっともらしい応答を生成できることにあるが、長い文脈・履歴を扱うことが難しく、対話全体を見通した当該発話の位置づけや、利用者の理解度合いの変化、話者間の知識共有のプロセスを扱うまでに至っていない。

一方で、鉄道会社の窓口業務での対話などにおいては、対話行為者には、ある程度想定された目

的と計画立案のモデルが適応でき、リットマンとアレン[4]をはじめとする、一連の計画認識の演算手法に関する研究によって、途中で計画に変更が加わる場合や、新たな計画が導入されるケースも含めて、かなり複雑な談話計画の遂行を計算機上で実現可能にする手法が考案されてきた。また、これらの対話を談話計画の遂行であると捉え、意図に基づく談話構造[5]を用いた、対話制御の理論的フレームワークも、古くから議論されてきたが、実際の自然談話に対する談話構造の定式化が難しく、そのままの形での応用は困難とされている。

本研究では、実際の旅行ガイドと旅行者の対話を対象に、彼らが、その対話によって、何を共有し、何を達成しているのか、その過程をモデル化することによって、ガイドの対話戦略モデルの記述を試みる。より具体的には、大局的な談話の目的(Discourse Purpose)、が、局所の談話セグメント[5][6]や発話交換[7]における、話者間の基盤化過程[8][9]の積み重ねで構築されると捉え、様々なレベルの基盤化ネットワーク[10]で記述することで、対話をモデル化することを試みる。基盤化ネットワークは、状態遷移モデルの一つであるので、対話戦略を、ネットワークとして計算機でそのまま扱える形式のモデルとなり得るが、表層・深層間の関係性や、理解・知識共有の程度を基盤化の強度に類する枠組みで扱えるモデルでなくてはならず、概念・フレームを拡張する必要がある。本発表では、その拡張に向けた取り組みと、実際の対話をモデル化する際につきあつた問題点について述べ、その原因と課題解決に向けた考察を行う。

#### 2. 方法

本章では、本研究のベースアイデアとなっている先行研究を紹介したうえで、本研究の対話のモデル化手法について述べる。

## 2.1 先行研究

グロスとサイドナー[5]は、目的のある談話への参加者がもつ、意図や目的の構造には階層性があり、それらが談話の言語構造にも反映されるとの考えのもと、意図に基づく談話構造理論を提案した。談話の言語構造は、発話や談話セグメント(discourse segment)と呼ばれる談話単位の列を階層的に組み上げたものと捉えられる。ここで、談話セグメントとは、対話参加者が、伝達すべき、あるいは、共有すべき目的(談話セグメント目的)を持った、ひとまとまりの発話の列である。例えば、電話で機器の起動の仕方を説明する、という対話においては、D1[機器を起動する]という談話目的が設定される。この談話目的に対して、例えば、D2[主電源を入れる]、D3[リモコンの電源を入れる]、という下位階層の談話セグメント目的があり、さらに、D2の談話セグメント目的を達成するためには、D4[主電源の場所を探す]、D5[主電源スイッチを押す]、という、より下位の談話セグメント目的を内包する、という具合である(談話セグメントの認定方法に関しては、[6]を参照のこと)。なお、談話セグメント目的の間には、支配(D1とD2、D3の内包関係)と充足先行(D4とD5の順序関係)の関係がある。グロスとサイドナーは、さらに談話参加者の注意状態の概念を導入し、彼らが、如何に談話構造を認識、参照しながら、談話目的を達成しようとしているかの、記述の枠組みを提案している。本研究における貢献トピック(後述)は、この談話セグメントに対応し、これをベースとするが、対象となる旅行案内対話の構造に基づいて、入れ子状の階層関係(支配関係)だけでなく、並列関係や複数の階層にまたがって共通性を持つまとまりについても許容するように設計している。

一方で、対話を、対話参加者間の共同活動であると捉える考え方がある。その活動を達成するためには、対話者間で様々な知識や信念が共有される必要があり、その共有過程を基盤化(Grounding)過程と呼ぶ[8][9]。音声発話をその基礎とする共同活動においては、話し手による一つ一つの発話が、受け手に理解された証拠が提示され、それを話し手も認識することによって、共有基盤を積み重ね、共同活動はその目的を達成していく。この、ある発話が受け手に理解されたことを示すまで発話の連鎖を、貢献(Contribution)と呼ぶ。貢献は、一組の発話交換であることもあれば、発話が理解される過程全体をさすこともある。貢献は、話し手による“提示”と受け手(および話し手による)“受理”の二段階で構成され、その受理は、受け手によ

る理解の証拠(あいづちや、応答など)により示される。もし、不理解の証拠(問い返しの「え?」など)が提示されれば、話し手は発話を再度繰り返すなどして修復し、理解の証拠が提示されることによって、一つの貢献が達成される。

トラウム[10]は、このアイデアを計算機モデルとして扱えるように拡張した。トラウムのモデルは、貢献に対応する談話ユニット(Discourse Unit)を規定し、この談話ユニットが、各発話の基盤化に対する行為ラベルである基盤化アクト(Grounding Act)の列によって構成されるというモデルである。基盤化アクトには、提示に対応するInitiate、受理に対応するAckをはじめ、同一主体による、内容追加に対するContinue、修復に対応するRepair、他者からのRepairの要求であるReqRepair、他者からのAckの要求であるReqAck、談話ユニットの処理を停止して、基盤化を無効にするCancelの、7種がある。さらにトラウムは、図1のように、現発話の基盤化アクトから、次の状態を決定するような、遷移ネットワークによって、基盤化過程をモデル化している。開始状態Sは、新しい談話状態が始まる前の状態である。このSから、開始者Iによって、Initiateが提示され、中間状態1に遷移する。さらに、同主体が、Continueによって、内容を追加し、受け手RのAckによって受理されて、最終状態Fに遷移し、この談話ユニットの基盤化が完了する。

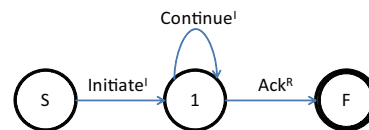


図1 基盤化の遷移ネットワーク例

また、他者開始の自己修復[12]などの修復連鎖を、中間状態からのサブネットワークとして表現することで、ネットワークに階層性の概念を導入し、さらに、サブネットワークの中間状態から再帰的に遷移を許容するような、再帰遷移ネットワークをモデル化している。本研究では、このトラウムのアイデアに基づき、次節のように拡張を試みる。

## 2.2 本研究の手法

トラウムのモデルでは、談話ユニット間にレベルの差はなく、前後関係は記述できても、ひとまとまりの談話ユニットで何を達成する貢献なのかを説明することを目的としていない。ゆえに、本研究では、ある談話ユニットのまとまりを、何を基盤化しようとしているのか、を定式化できるような、談話セグメントに対応する“貢献トピック”と

して定義し、かつ、基盤化のレベルを、内部状態として持つように状態を拡張し、対話全体を階層的にモデル化する。ここにおいて、基盤化のレベルとは、トラウムのモデルにおける中間状態を拡張し、ロックとトラウム[13]のアイデアを元に、不十分な基盤化状態と、修復を要する基盤化の失敗状態のような、基盤化以前の、対話参加者の共有対象の理解状態をレベル分けしたものである。また、貢献トピックと、談話セグメントの違いは、発話列のどこから、どこまでを一つの貢献とみなすのか、の基準を、共有すべき命題（「旅行者の要求はXである」「Xに対する回答はYである」など）が、対話者間で基盤化完了に至ったかどうか、に置く点、および、トピックを対象ドメイン（本研究においては電話観光案内）において、ある程度共通性をもってラベリングし、レベル（階層）分け可能なようにデザインする点にある。また、一つの発話が、複数の命題を内包する場合には、それを並列して、ラベルすることを許容した。ただし、この点に関しては、基盤化ネットワークとの不連続性に問題があり、5.において、考察する。また、基盤化アクトに関してはトラウムの7つ（Initiate, Continue, Ack, Repair, ReqRepair, ReqAck, Cancel）をほぼそのまま用いたが、複数の談話ユニットをつなげて表現するために、Ack/Initiate, という便宜的な基盤化アクトを加え、間に受け手の相槌（Ack）を挟んだ、開始者のInitiateの続きとなる発話に関しても、Continueを付与した。また、ReqAckの解釈を、単なる承認の要求だけでなく、（自分の認識の）確認のためのYes/No質問もこれに含めた。

### 3. データおよびラベリング

本研究では、ガイドと旅行者の対話を対象とする。旅行者は、GPS機能付きの携帯電話を持ち、実際に京都をめぐりながら、現地で生じた質問、相談をガイドとの電話対話を通じて解消するというもので、電話の呼び出しから、電話を切る（あるいは切れる）までの一連の対話が、一つの対話セッションである。ガイドは、旅行者の大まかな位置は取得できるが、リアルタイムの位置や方角は把握できないので、それを確認しながら、対応することになる。特に道案内に関しては、ある程度まで説明して、一旦電話を切り、再度電話をもらってから、続きを説明する、ということもあった。一人のプロのガイドが、一日5名、計10名の旅行者の一日旅行に対する案内を受け付けた。本発表では、このうち、一人の旅行者による、計7回の電話依頼対話データ、および、比較対象として、異なる旅行者の計6回のデータも分析対象とした。

これらのデータに対し、ラベル設計者が、マニュ

アルを整備したうえで、サンプル対話を例に、一人の作業者に指導し、対象データに付与した。別途ラベル設計者もラベルを付与し、不一致点に関しては適宜協議のうえ、ラベルを確定した。

### 4. 結果

上記のデータに対して、それぞれのラベルを付与した結果の例が、図2である。一行目のラベル名の番号は、最初の番号が、当該トピックのレベル（階層）、後の番号が、同階層で並列な貢献を表す。この例の場合は、第一階層と第二階層があり、並列・重複が、第一階層で二つ、第二階層で三つ存在する。談話全体を表す第0階層、第一階層は、下位の貢献トピックでのみ構成されるが、第二階層以下（第三、第四も別の例で存在する）は、基本的に、最下層の貢献トピックに、基盤化アクト・状態ラベルが付与される。ただし、発話の内容に依存して、複数階層に、基盤化アクト・レベルが付与される場合がある。貢献トピックの種類は、第一階層で5種、第二階層で6種、複数の階層に付与し得るものが6種の、全部で17種となった。

対象データ全てに付与し、それらを一つのネットワークに統合した結果、この旅行者とガイドの対話は、トップレベルの貢献トピックの遷移（第0階層の遷移ネットワーク）として、図3のような状態遷移ネットワークで記述された（Sが開始、Fが終了状態）。中間状態の数は、第一階層の貢献トピックの出現順序と7つの対話データの共通性に依存し、この例の場合、2であった。図2に示したデータ例のケースが、最も単純なケースで、最初のやり取り<opening>の後、旅行者の依頼対応<request>があり、電話を切るためのやり取り<closing>が行われて会話は終了する。少なくとも、このドメインにおいては、この遷移が基礎となる。場合によって、一つの<request>の後、再度別の<request>があったり（状態2からの自己遷移）、ガイド側から、自発的な提案<proactive>があったり（同）、本題からそれるが、ガイドー旅行者間の良好な関係を維持するための会話<addition>がある（同）こともあるが、<closing>によって、最終状態に至る（まれに電波のトラブルで、対話が基盤化されずに途切れることもある）。なお、最終状態から状態2への遷移は、一旦、会話を終えるための貢献が完了した後、ガイドが天気の話を加えたためであるが、再度、<closing>によって、その対話は完了した。

これら第一階層の貢献トピックは、また第二階層の貢献トピックによるサブネットワークで表現される。一例（<request>）を、図3に示す。基本的な遷移としては、質問の前提となる先の目的<goal>（「最終的には金閣寺に行こうと思ってい

Role	Utterance	ID	Topics1-1	Topics1-2	Topics2-1	Act2-1	CG2-1	Topics2-2	Act2-2	CG2-2	Topics2-3	Act2-3	CG2-3
	<Call>	1	<<opening>>		<channel>	initiate	ui						
g	はい、山田です。	2				ack	g	<identify>	initiate	gi			
u	田中ですけれども。	3				ack	u+		ack/initiate	u+			
g	お疲れさまです。	4			</channel>	ack/	g+m/	</identify>	ack/	g+m/	<social>	initiate	gi
u	たびたびすみません。	5									</social>	ack/initiate	u+
g	いえ。	6	<<opening>>										
u	すみません、清水寺のなかなんですけれども。	7	<<request>>		<situation>	initiate	ui						
g	はい。	8			</situation>	ack/	g+m/						
u	今、おみくじを引いてまして、	9			<undreq>	initiate	ui						
g	hhhh	10				ack	g						
u	hhhhちよつと忘れちゃったんですけど、いいほうを結ぶんだっか、悪いほうを結ぶんだっか。	11					continue	uc					
g	悪いほうを結んで下さい。hhhh	12			</undreq>	ack/	g+m/	<answer>	initiate	gi			
u	悪いほうを結んで、いいのは持つんでしたっけ？	13							reqack	u			
g	そうです。そうです。	14							ack	g+			
u	hhすみません。ありがとうございますまよかったですか？	15	<<request>>	<<closing>>	<social>	initiate	ui	</answer>	ack/	u+m/			
g	はい、良かったです。hhh	16			</social>	ack/	g+m/	<addition>	initiate	gi			
u	はい、良かったです。hhh	17							ack/initiate	u			
g	hhhお持ち帰りください。	18							ack/initiate	g+			
u	はい、ありがとうございます。	19			<social>	initiate	ui	</addition>	ack/	u+m/			
g	はい、ありがとうございます。	20				ack	g+						
u	はい、失礼します。	21			</social>	ack/	u+m/	<social>	initiate	ui			
g	失礼します。	22		<<closing>>				</social>	ack/	g+m/			

図 2 ラベルされた対話データ例

るんですが」など)か、あるいは、今現在の状況説明<situation> (「今、清水寺の門のところにいるんですけど」など)があり、その後、依頼の内容に関わる発話<und:req>があり、それに対するガイドの回答<answer>があって、その基盤化が完了すれば、最終状態に至る。<pre:req>は、「質問してもよろしいですか?」のような、依頼に入る前の断りの連鎖がなされることがあり、ほぼ初回の電話のときだけに見られるので、初期状態からの自己遷移としている。なお、初期状態からすぐに依頼内容を語りだすケースもあった(初期状態から状態2への遷移)。

さらに、第二階層の貢献トピックの基盤化例を、図5に示す。これは、旅行者の依頼を理解した上で、ガイドがその依頼に答え、旅行者がそれを理解する<answer>の基盤化過程である。遷移の中には、<situation>という貢献トピックも存在し、さらに下位の階層の遷移も含まれている。遷移は、必ず、Initiateから始まるので、まず、gi (あるいはui) に遷移する。例は、回答なので、ほぼ、ガイドが開始者となる。u (あるいはg) は、まだ基盤化途中の、「そこまではとりあえず理解できたことが互いに基盤化された」状態であり、ContinueやAck, ReqAckを提示しながら、この中間状態への遷移を繰り返し、命題(この場合はガイドの回答の内容全て)が受け手に理解された証拠が提示され、それが相互に共有された状態(ガイドの回答が旅行者に理解され、それが理解されたことがガイドにも理解された状態)、つまりは、最終状態mに至る。u- (あるいはg-) は、不理解の証拠が提示された場合(ほとんどの場合は、Reqrepair修復における他者開始)の状態である。このほか、u! (あるいはg!) は、命題の意味は理解したが、それ

に答えられる状態になかったり、理解に自信がないという証拠が提示された結果、基盤化に貢献しない状態(例えば、「どこにいますか?」という問いに対して、「どこかわからない」と回答した場合など)とし、特に、<situation>内の遷移ネットワークに見られた。

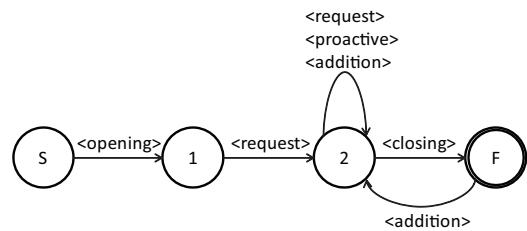


図 3 電話旅行案内対話の遷移ネットワーク(第0階層)

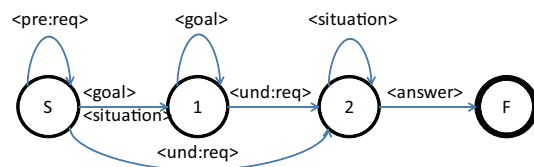


図 4 <request>の遷移ネットワーク(第一階層)

### 5. 考察

本研究における、貢献トピックの構造は、多くの談話構造研究における、中位の構造化分類手法、例えば、「話段[14]」に類似することは明らかである。ただし、前述のように、貢献トピックは、基盤化の過程に基づいて設計されているという点と、対話者(ガイド)の戦略をモデル化するこ

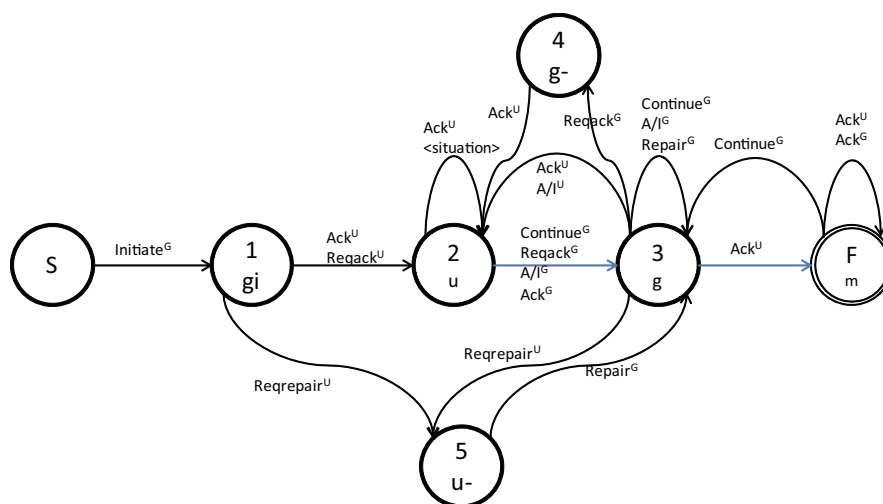


図5 &lt;answer&gt;の遷移ネットワーク（第二階層）

とを意図しているために、各貢献の目的をある程度明確に、そしてある程度般化できるような粒度で、ラベル設計をしている点が異なると言える。なお、多くの話段による談話構造の分節においては、<request>を、旅行者による問いかけと、ガイドによる回答部分にわけると一般的であるが、今回のデータからは、<situation>などとともに、<request>内のサブネットワークとして構成したほうが、ラベルの粒度として適切であると判断した。

第0階層や、一部の第一階層（<opening>や<closing>など）の遷移ネットワークは、電話による旅行案内対話においては、ほぼ普遍的であると思われるが、第二階層以下においては、状態数や遷移が流動的で、普遍的とは言い難い。ただし、これはガイドの戦略と言うよりも、その時々での偶発的な遷移によって生じたものも多い。そのため、異なる旅行者のデータに対し、同様に付与した結果、<request>のネットワーク構造はほぼかわらなかったが、第二階層以下の貢献トピック、例えば<answer>の構造は、大きく異なるものとなった。特に、回答中の状況確認（<situation>）や目的の確認（<goal>）、回答のための深堀の質問（<probe>）など、第三階層のサブネットワークへの遷移を含む対話が多く、より複雑なものとなった。旅行者の質問・依頼は、図2に示したような、単純な質問／応答として位置づけられるものから、道案内のような、最終的な目的を達成するために、数多くの確認作業が必要になるものまで、その対話継続長を含めて多種多様である。極論を言えば、全ての対話を一つの遷移ネットワークに落とし込もうとすれば、結局、最大の状態数に対して、全ての基盤化アクトによる遷移が加えられた、大きなネッ

トワークができるだけであり、何らモデルは単純化されない。例えば、これらの遷移を頻度などで制約することで、モデルを簡略化することは可能である。しかし、それら一回的な遷移を削除することは、対話の自然さを欠くことにもつながるため、制約の適切な調整が必要となるだろう。

また、貢献トピックの変遷を、遷移ネットワークに変換するために問題となったのは、図2の<opening>に見られるような、同じ階層に、並列したトピックがある場合のネットワーク表現である。<opening>は、電話回線の確保<channel>、互いが誰であるかの確認<identify>、社会的（儀礼的）やり取り<social>の各基盤化過程からなり、一つの発話が、複数の貢献トピックに関わる。これを表現するには、一つのネットワークだけではなく、並列した別のネットワーク（サブネットワークではなく）を用意しなくてはならない。一方のネットワークに会話が遷移している間は、もう一方のネットワークは一旦停止状態とし、先のネットワークが完了した時点で、改めて遷移を開始する、というモデルであれば、実際の計算機モデルとしても可能であるが、同時に遷移を動かすようなモデルを想定することは難しく、今後工夫を要する。あるいは、ラベル手法そのものを見直す必要がある。

今後、旅行者10名分の対象データ全てのネットワークを記述した上で、旅行対話における対話モデルの普遍的遷移と、必ずしも必要でない偶発的な遷移との差異化、あるいは対話システムのための効率化と、対話の自然さの確保とのバランスを保つための方法論について、考察を進めていきたいと考えている。なお、本研究の一部は、科研費基盤研究(C)（課題番号24520447）の助成を受けて



行われた。

## 参考文献

- [1] Hori, C., Ohtake, K., Misu, T., Kashioka, H. and Nakamura, T., (2009). “Weighted Finite State Transducer Based Statistical Dialog Management”, Proceedings of Automatic Speech Recognition and Understanding, 2009IEEE, pp. 490-495.
- [2] 目黒豊美・東中竜一郎・南泰浩・堂坂浩. (2011). “POMDPを用いた聞き役対話システムの対話制御”, 言語処理学会 第17回年次大会 発表論文集, 913-915.
- [3] Boyer, K. E., Ha, E. Y., Wallis, M. D., Philips, R., Vouk, M. A. and Lester, J. C. (2009). “Discovering Tutorial Dialogue Strategies with Hidden Markov Models”, Proceedings of the Fourteenth International Conference on Artificial Intelligence in Education, 141-148.
- [4] Litman, D. J. and Allen, J. F. (1987). “A plan recognition model for subdialogues in conversations”, Cognitive Science, 11, pp.163-200.
- [5] Grosz, B. J. and Sidner, C. L. (1986). “Attention, intention, and the structure of discourse”, Computational Linguistics, 12, 175-204.
- [6] Christine H. Nakatani. Barbara J. Grosz. David D. Ahn. Julia Hirschberg. (1995). “Instructions for Annotating Discourse”, TR-21-95.
- [7] Coulthard, R. M. (1985). “An introduction to discourse analysis (2nd ed.)”, Longman.
- [8] Clark, H. H. and Schaefer, E. F. (1989). “Contributing to discourse”, Cognitive Science, 13, 259-294.
- [9] Clark, H. H. and Brennan, S. A. (1991). “Grounding in communication”, In: Resnick LB, Levine JM, Teasley SD (eds) Perspectives on socially shared cognition, APA Books, Washington, pp 127-149
- [10] Traum, D. (1994) “A computational theory of grounding in natural language conversation”, Unpublished doctoral dissertation, University of Rochester.
- [11] Traum D. (1999) ‘Computational models of grounding in collaborative systems’, working notes of AAAI Fall Symposium on Psychological Models of Communication, November, pp.124-131.
- [12] Schegloff, E. A. and Jefferson, G. and Sacks, H. (1977) “The preference for self-correction in the organization of repair in conversation”, Language, 53, pp.54-63.
- [13] Roque A, Traum D (2008) “Degrees of grounding based on evidence of understanding”, In: proceedings of The 9th SIGdial Workshop on Discourse and Dialogue, pp.54-63.
- [14] ポリー・ザトラウスキー, (1991). “会話分析における単位について-「話段」の提案”, 日本語学, 10:10, pp.79-96.