

Title	対話のセグメンテーションに関する研究
Author(s)	金, 美慶
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/665
Rights	
Description	Supervisor:石崎 雅人, 知識科学研究科, 修士

修 士 論 文

対話のセグメンテーション方式に関する研究

指導教官 石崎雅人 助教授

北陸先端科学技術大学院大学
知識科学研究科知識システム基礎学専攻

850029 金 美慶

審査委員： 石崎 雅人 助教授（主査）

Ho Tu Bao 教授

下嶋 篤 助教授

2000 年 2 月

目 次

1	はじめに	1
2	関連研究	3
2. 1	談話セグメント	3
2. 2	隣接ペア、交換構造	5
3	対話セグメンテーション方式の提案と評価	8
3. 1	利用するデータ	8
3. 2	対話セグメンテーションの方式の 提案と評価	10
3. 3	考察	17
4	セグメントの意味的側面	
	ー主導権の概念からー	22
4. 1	主導権の定義	22
4. 2	主導権の観点からのセグメント	27
5	おわりに	30

目 次

1	映画に関するエッセイの例	4
2	談話分析に基づく談話構造	6
3	「大学案内依頼」の談話構造の例	7
4	やりとりの例	9
5	談話セグメントタグ	9
6	話者ごとの発話	10
7	マーキングをする発話のリスト	11
8	再現率と精度の関係グラフ	12
9	繰り返しの例	13
10	質問－応答の例	14
11	セグメント境界の外にある発話に対する 応答の例	16
12	質問－（応答－了解） ⁿ の例	17
13	質問－応答－情報提供－受諾の例	17
14	質問－応答（繰り返しを含む）の例	19
15	質問－応答－繰り返しの例	20
16	質問－応答／質問－応答の例	20
17	質問－応答（ステージング）の例	20
18	質問－応答*の例	21
19	主導権分析のための例文1	24
20	主導権分析のための例文2	26
21	主導権の分析	27
22	主導権分析のための実例	28

表 目 次

1	方式 1 の再現率と精度	1 3
2	方式 1 の平均と標準偏差	1 3
3	方式 2 の再現率と精度	1 5
4	方式 2 の平均と標準偏差	1 5
5	手がかり語を利用した再現率と精度	1 8
6	手がかり語の平均と標準偏差	1 8

第 1 章

はじめに

現代人の我々は、技術発展に伴い様々なメディアを用いてコミュニケーションするようになった。そのようなコミュニケーションには、人間同士だけではなく、我々の生活の中に深く浸透している機械とのコミュニケーションも含まれる。人間と機械のコミュニケーションは、人間と人間とのコミュニケーション（対話）に比べると貧弱で、正確に言うところコミュニケーションが成立しているとは言いがたく、多くの研究者たちが効率よく自然に機械とインタラクションできるシステムをつくるために研究を進めている。人間と人間のコミュニケーション（対話）は、必ずしも人間と機械のコミュニケーションと同じではないが、そのメカニズムを解明することにより、人間と機械のコミュニケーションを改善するための知見を得ることができるのではないかと考える。

本論文では、音声対話コーパスを利用して、対話の分析の単位となるセグメントを精度よく取り出す方法の提案を行い、また、抽出した単位について主導権の観点からの分析を行う。分析の対象としては、できるだけ多様な対話データを利用することにした。多様といっても、雑談のような構造を判断するのが極めて難しいものは避け、音声言語研究で主に利用されている課題遂行対話を扱った。多様な対話とは、すなわち課題のバリエーションのある対話を意味する。

そもそもなぜセグメントを決めることが重要なのであろうか。文章や独話の場合に、その目的を達成するまとまりがあり、そのまとまりは韻律的、言語的特徴をもっていることが明らかにされ、例えば英語の場合、照応関係の決定に利用できることが示されている。このようなまとまりを正しく見つけることができれば、理解する対象を、文章の独話全体から、より小さいまとまりにすることができるという意味で計算機で処理がしやすくなる。また、このようなまとまりごとに理解するというやり方は、（非

常に直感的ではあるが)我々が利用している方法と言ってもよいであろう。

さて、本論文で対象とする課題遂行対話の場合はどうであろうか。対話は課題を遂行するためのステップを構成するまとまりから作り上げられるという意味では文章、独話と同じであり、まとまりを正しく認識することは文章、独話と同じ利点をもつ。

現在までの研究ではどのようなセグメントを用いて対話分析を行ってきたのだろうか。精度よく、かつ、簡単にセグメントを決定することができれば計算機による対話理解の精度を高めたり、対話分析のために大量のデータを作成するのに役立つであろう。セグメントの候補として、計算言語学の分野で提案されている談話セグメント (**discourse segment**)、社会学の一分野である会話分析の隣接ペア (**adjacency pair**)、言語学の一分野である談話分析の交換構造 (**exchange structure**) がある。談話セグメントについては、対話についてその決め方が明確でなく、人によって異なる結果になるため、正確なデータを作成するのが難しいという問題がある。また、交換構造と隣接ペアについては、それらが発話内行為または談話行為に基礎を置いているが、これらの分類は高い精度で行うのが難しいという問題がある。

仮にセグメントに区切ることができたとして、このようなセグメントはどのような意味を持つのだろうか。対話の単位の意味的な側面を理解する概念として主導権という概念が提案されている。そこで本論文では、提案した基準で区切ったセグメントを、さらに主導権という概念から分析する。以下、2 章では、従来で利用されている単位 (セグメント) について説明する。3 章では、今回利用したデータについての説明、本研究での方法で得られた結果と分析から、その有用性を明らかにする。4 章では、本研究で得られたセグメントの意味的な側面について、主導権という概念を利用して議論を行う。

第 2 章

関連研究

1 章では、対話の分析の単位を取り出すための対話セグメンテーションに利用できる考え方として、

- ・ 談話セグメント
- ・ 質問／応答などのやりとりの構造（隣接ペア、交換構造）

があることについて簡単に触れた。本章では、これらの考え方を詳細に検討し、問題点を明らかにする。

2.1 談話セグメント

（Grosz and Sidner, 1986）は、談話を理解するためには、言語、意図（intention）、注意（attention）の 3 種類の情報の相互関係を考えることが必要であると主張した。言語情報は、韻律、語彙、構文情報を含み、意図情報とは、談話参加者の目標状態を指す。注意情報とは、談話の進展に伴う談話参加者の注意のあり方をモデル化したものである。言語情報は、意図の情報がどのようになっているかを示唆したり、注意情報は、例えば、照応関係についての制約を課すといった具合にこの 3 種類の情報は相互に関連している。映画に関するエッセイの例（図 1）を使って実際に見てみる。

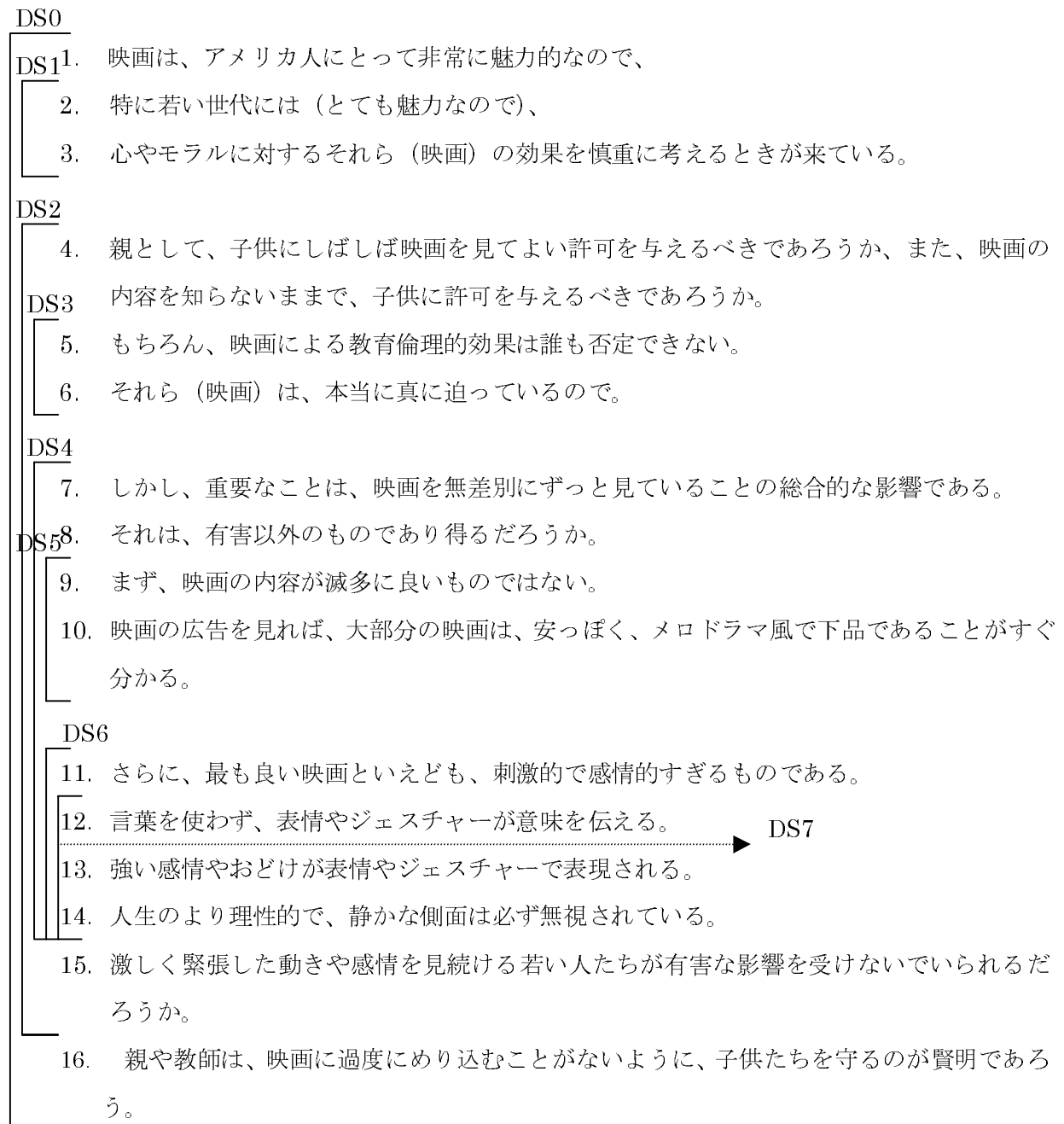


図 1 映画に関するエッセイの例 (Grosz and Sidner, 1986)

図 1 で、意図情報で境界づけられるセグメントは、DS0、DS1、DS2、…のように表されている。DS0、すなわちこのエッセイ全体の意図は、親と教師が、子供たちが過度に映画にのめりこまないように守るべきであるということを伝えることである。

この DS0 は、さらに心やモラルに対する映画の影響を考えると来ているという問題提起 DS1 と激しく緊張した動きや感情を見続ける若い人たちには有害な影響があるという主張 DS2 を含んでいる。

言語情報は、例えば DS1 から DS2 へ移るときには、平叙文から疑問文へのアスペクトの変化や DS5 がはじまるときのように「まず」といった手がかり語の出現が意図構造と相関して現れている。

注意情報は、セグメント中に現れる個体、属性、関係を含むので‘映画’、‘アメリカ人’に‘映画はアメリカ人にとって非常に魅力的’といったものを含む。

この理論自体は文章、対話の区別をせず話しをすすめているが、実際は(Nakatani et al., 1995)にあるように、実際のデータを扱うためにまず独話を対象にしている。彼女たちは、発話の動機を作業者に推測させるというきわめて簡単な方法で精度よくセグメント境界を決めることができると主張している。なぜ独話を対象にしたかについて彼女たち自身からの釈明はないか。対話の場合は、意図構造が交錯することが予想され、彼女たちの(Nakatani et al., 1995)で採用した単純に発話動機を明示する方法では扱うのが難しかったのではないかと考える。

2.2 隣接ペア、交換構造

会話分析(Schegloff and Sacks, 1973; Levinson, 1983)は、対話の単位として質問—応答、依頼—受諾／拒否のような隣接ペアがあることを指摘している。この理論は、社会学の理論として、このような会話の分析から社会の諸現象がいかに創り出されていくかを明らかにするために利用されている。

談話分析(Coulthard, 1985; Stenström, 1994)では、会話分析、言語行為論(Austin, 1962; Searle, 1969)、システムック文法(Halliday, 1976)をもとに、対話(談話)の構造を分析する理論を提案している。

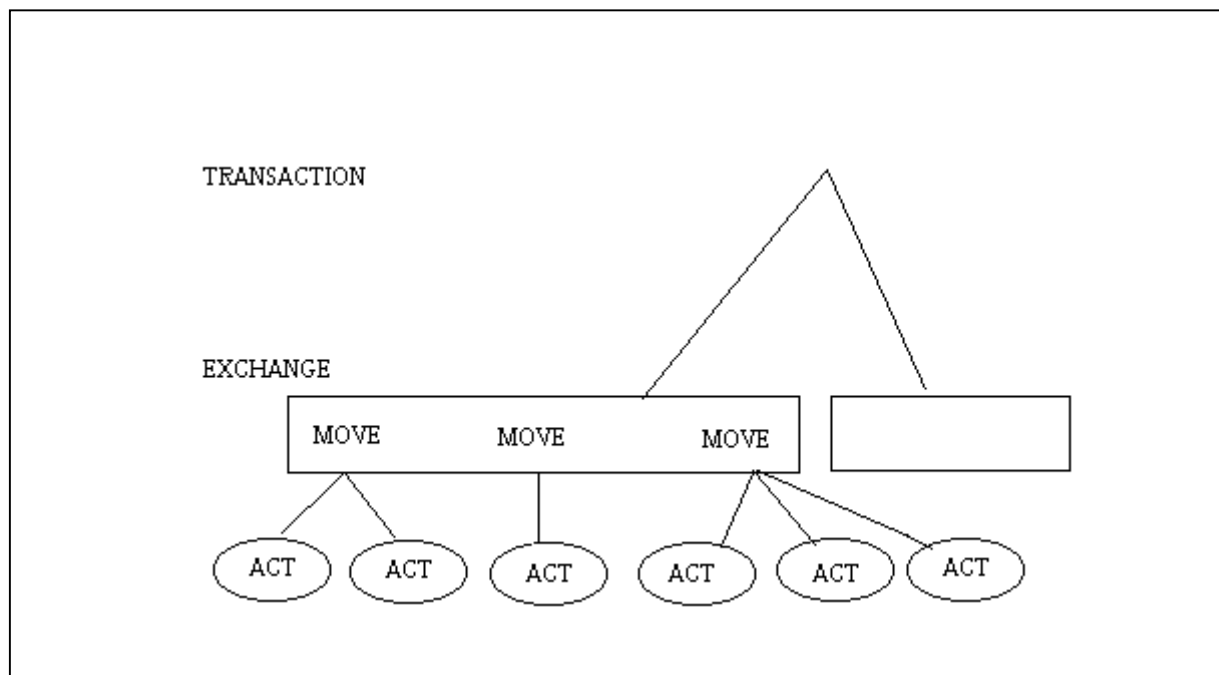


図 2 談話分析に基づく談話構造

談話分析では、談話は **Act**（談話行為）、**Exchange**（交換構造）、**Transaction**（話題構造）の 3 つの層から構成されるとする（図 2）。**Act** は質問、依頼、応答のような談話行為である。**Act** は単独、あるいはいくつかが集まり、**Move**（交換単位）を構成する。**Move** は、単独あるいは複数が集まって **Exchange** を構成する。**Exchange** の中で **Move** は働き掛け（**Initiate**）、応答（**Response**）、了解（**Followup**）の役割をもち、働き掛け **Move** によって開始し、その発話の示した目的が達成あるいは破棄されることによって終了する。**Transaction** は対話中の複数の **Exchange** から構成され、あるレベルの話題について述べるものである。本論文では、談話のセグメント化が主眼なので、構造的側面に着目して談話分析に焦点を絞る。「大学案内依頼の例」における構造を図 3 に示す。

ここで大括弧に囲まれた部分が **Transaction** である。図 3 で分かるように、発話 3～9 は、3 つの **Exchange** から構成されている。発話 1～2 と 10～13 の **Exchange** はそのまま **Transaction** であることが分かる。

談話分析に基づいて対話をセグメント化しようとする場合、まず **Act** が基本になる。しかし、**Act** そのものの認識はそれほど簡単な作業ではなく、**Act** に推定する発

話を規定するのが難しい（荒木他, 1999）。

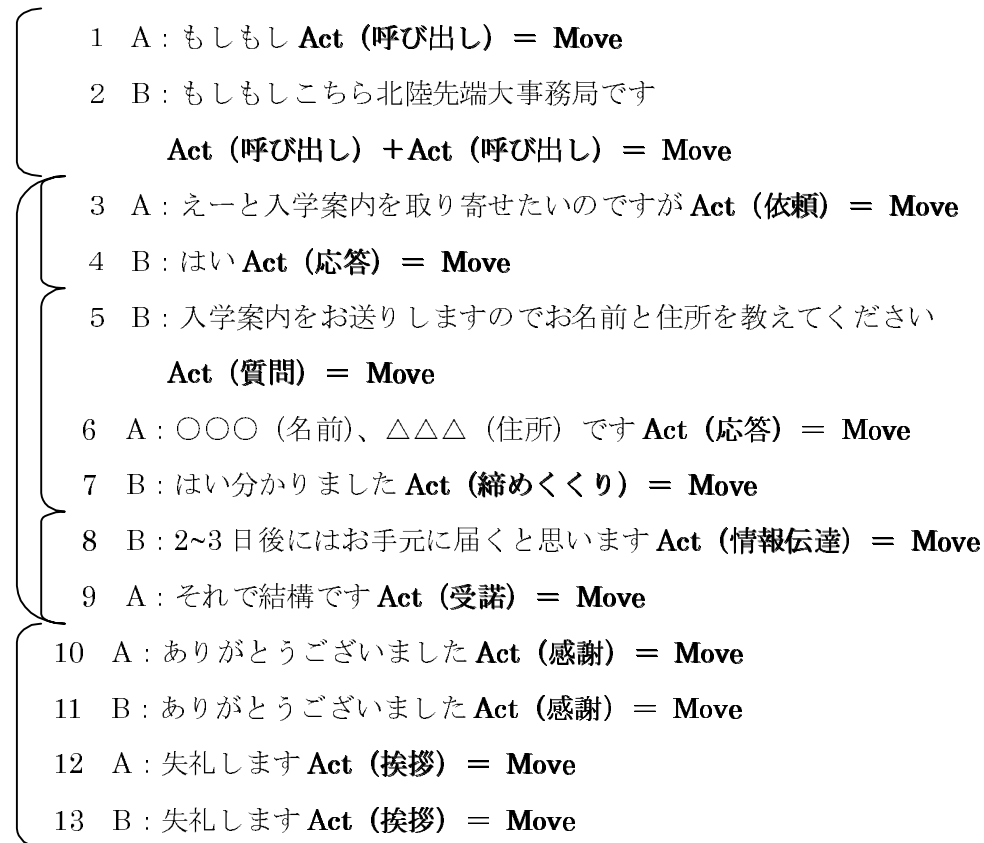
- 
- 1 A : もしもし **Act (呼び出し) = Move**
- 2 B : もしもしこちら北陸先端大事務局です
- Act (呼び出し) + Act (呼び出し) = Move**
- 3 A : えーと入学案内を取り寄せたいのですが **Act (依頼) = Move**
- 4 B : はい **Act (応答) = Move**
- 5 B : 入学案内をお送りしますのでお名前と住所を教えてください
- Act (質問) = Move**
- 6 A : ○○○ (名前)、△△△ (住所) です **Act (応答) = Move**
- 7 B : はい分かりました **Act (締めくくり) = Move**
- 8 B : 2~3 日後にはお手元に届くと思います **Act (情報伝達) = Move**
- 9 A : それで結構です **Act (受諾) = Move**
- 10 A : ありがとうございます **Act (感謝) = Move**
- 11 B : ありがとうございます **Act (感謝) = Move**
- 12 A : 失礼します **Act (挨拶) = Move**
- 13 B : 失礼します **Act (挨拶) = Move**

図3 「大学案内依頼」の談話構造の例

第 3 章

対話セグメンテーション方式の

提案と評価

3.1 利用するデータ

本論文では、千葉大学、早稲田大学、大阪大学、筑波大学、京都大学、電気通信大学、ATR 音声翻案通信研究所、通信総合研究所で収録され、人工知能学会、言語音声理解と対話処理研究会、談話・対話研究におけるコーパス利用研究グループによってさまざまな情報が付与された総 14 対話を利用する（コーパス利用 WG, 2000）。以下にデータ量、課題、付与されている情報をまとめる。

- ・データ量：総計 52 分 40 秒
- ・課題：地理課題、クロスワードパズル、会議の予約、地理案内、テレフォンショッピング、スケジューリング、ホテルの予約、父親当てクイズ
- ・付与されている情報：
 - 時間：単位の開始・終了時刻を付与
 - かな漢字による転記：発話は 400ms で区切ったスラッシュ単位として再構成
このとき談話標識などをマークしている
 - 形態素：形態素解析システム『茶釜』を用いて単語、品詞などの情報を付与
 - 談話行為タグ
 - 談話セグメント
 - 韻律

談話行為タグは、対話の基本的構成が「対話開始」「対話終了」「問題解決」の部分

からなると定義し、中心となる問題解決部は1つ以上のやりとりからなるとする。やりとりは問題解決部の基本単位で、基本的には「働きかけ (Initiate)」 - 「応答 (Response)」 - 「了解 (Followup)」からなる。基本的なやりとりの例を示す。

.....

(I) 0041 A : {F え} {F あの} 地下鉄の駅名は
(R) 0042 B : て {F え} 地下鉄の寺町駅になります
(F) 0043 A : あはい

.....

I : 働きかけ、R : 応答、F : 了解

図4 やりとりの例

働きかけタイプには「示唆」、「依頼」、「提案」、「勧誘」、「確認」、「真偽情報要求」、「未知情報要求」、「約束・申し出」、「希望」、「情報伝達」、「その他の言明」、「その他の働きかけ」があり、談話行為タグとして付与される。応答タイプは「肯定・受諾」、「否定・拒否」、「未知情報応答」、「その他の応答」が、了解タイプはやりとりをしめくくる「了解」がある。

談話セグメントは、働きかけタイプの談話行為を持つ発話の前に談話セグメントが存在すると考え、談話セグメント境界（タグ）を置く。話題セグメントタグは、話題境界らしさ（1,2の2段階）、話題名、セグメント関係情報で構成される。セグメント関係情報には「聞き返し」、「割り込み」、「復帰」がある。

.....

[1 : 大会議室 :]

0041 A : 大会議室が使えます

[1 : 講演回の場合 : 復帰]

0042 B : {D じゃ} 大会議室でお願いします

.....

図5 談話セグメントタグ

今回利用したデータは、発話レベルで考えて統計的処理を前提する場合だとデータ数は多いとは言えない。しかし、対話の課題にバリエーションがあり、各対話に豊富

な情報が付与されているという特徴がある。逆に言えば、豊富な情報が付与されたデータを多量にアクセスすることが難しいという現状があるので、本論文では課題が多種類であることと情報の多さから利用することにした。

3.2 対話セグメンテーションの方式の提案と評価

対話は2人以上の相互行為である。対話を見るとき、直感的ではあるが発話は断片のように見える。ところが、一人一人の発話の流れをみると、断片ではなく、それなりのまとまりがある。そしてそのまとまりは、相手の応答への「はい」などの実質的には意味をもたない発話で区切られている。

x	0000	L:まず	副詞・一般	0000	R:はい	感動詞	
x	0001	L:出発	名詞・サ変接続	0001	R:はい	感動詞	
x	0002	L:地点	名詞・一般	0002	R:うん	感動詞	
x	0003	L:から	助詞・格助詞・一般	x	0003	R:ひ	語断片
x	0004	L:西	名詞・一般	x	0004	R:左	名詞・一般
x	0005	L:側	名詞・接尾・一般	x	0005	R:で	助動詞
x	0006	L:に	助詞・格助詞・一般	x	0006	R:いい	形容詞・非自立
x	0007	L:すこし	副詞・助詞類接続	x	0007	R:ん	名詞・非自立・一般
x	0008	L:だけ	助詞・副助詞	x	0008	R:だ	助動詞
x	0009	L:行き	動詞・自立	x	0009	R:よ	助詞・終助詞
x	0010	L:ます	助動詞	x	0010	R:ね	助詞・終助詞
	0011	L:はい	感動詞	0011	R:うん	感動詞	
x	0012	L:で	接続詞	0012	R:うん	感動詞	
x	0013	L:大滝	名詞・一般	0013	R:うん	感動詞	
x	0014	L:が	助詞・格助詞・一般	0014	R:ある	動詞・自立	
x	0015	L:ある	動詞・自立	0015	R:ある	動詞・自立	

図6 話者ごとの発話

図 6 の左右はそれぞれ別のファイルになっている。このデータには、情報として単語番号、話者、単語、品詞が付与されている。一人ずつの発話の流れを見て、実質的意味をもたないと考えられる発話とそうでない発話を区別するための作業を行った。形態素が付与されたファイルに **x** をつけ、実質的意味のない単語だと思われる箇所だけの **x** を消した。方式 1 で説明されるマークとは **x** を消すことである。

左の発話だけをみて、「はい」で区切られる前のまとまりは“出発地点からの移動”を、後のまとまりは“大滝”の話しであると理解できる。この考え方を基に談話セグメンテーションができると考え、方式 1 を提案する。

方式 1

一人一人の発話の流れを見て「はい」、「分かりました」などの実質的に意味をもたない単語をマークし、マークのない発話からセグメントがはじまり、次にマークのない発話が現われる間をセグメントとする。図 7 にマークする単語のリストを示す。

「はい」「いいえ」「ある」「ない（か）」「わかりました」「うん」
「そう（そうそう）」「ありません」「ないです」「ああああ」
「そうです（か）」「いや」「はあ（はあはあ）」「は」「それ」
「そうございますか」「ありがとうございます」「たぶん」
「とんでもありません」「お願いします」「ふん（ふんふん）」
「そういうことなります」「よろしいです」「なるほど」
「かしこまりました」「もうね」

図 7 マーキングをする発話のリスト

方式 1 の評価

対話データに付与されている談話セグメントを正解と考えて再現率、精度で評価する。再現率は、正解セグメント数 (A) の中にどのくらい正解セグメントの中から提案する方式で見つけることができたセグメント数 (B) があるかをパーセンテージで示した値である。

$$\text{再現率 (\%)} = \frac{B}{A} \times 100$$

精度は、提案する方式で認識されたセグメント数 (C) において、提案方式で見つけた正解セグメント数 (B) の割合である。

$$\text{精度 (\%)} = \frac{B}{C} \times 100$$

再現率と精度は、正解セグメント (A) 集合と方式 1 で発見されたセグメント (C) 集合の積集合 (B)、すなわち一致したセグメント集合で表現することができ、それぞれの集合に占める積集合の割合である。

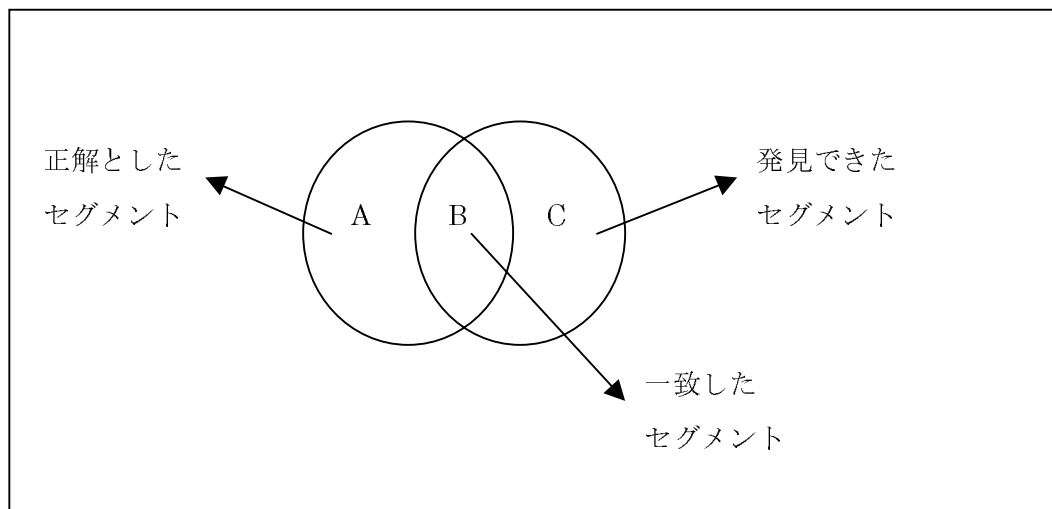


図 8 再現率と精度の関係グラフ

表 1 には方式 1 で得られたセグメントの再現率と精度を示す。

対話	再現率(%)	精度(%)	対話	再現率(%)	精度(%)
99atr01.ds1	100.0	62.1	99kyo0a.ds1	100.0	75.8
99atr02.ds1	100.0	78.6	99osa01.ds1	100.0	54.2
99chi0a.ds1	98.0	87.3	99osa02.ds1	98.5	65.0
99crl01.ds1	95.0	73.1	99osa0a.ds1	95.7	57.9
99crl02.ds1	96.6	73.7	99tsu0a.ds1	100.0	45.6
99crl03.ds1	97.1	85.0	99uec0a.ds1	95.8	62.2
99kyo01.ds1	100.0	64.0	99was03.ds1	100.0	78.0

表 1 方式 1 の再現率と精度

	再現率	精度
平均	98.3	68.7
標準偏差	1.94	12.0

表 2 方式 1 の平均と標準偏差

評価結果

方式 1 は、再現率が高いが、精度が低いことが分かる。誤って判断するところを分析した結果、そこには 2 つの特徴が見られた。

①同じ表現（単語、句）の繰り返し

②質問－応答、依頼－受諾、開始と終了の挨拶などの隣接ペア

図 9 と図 10 はそれぞれ①と②の例である。

[1：魔術山の周囲での移動]

x 0093 L：

*その左側を*こう通って

0094 R：

* {B うん}

o 0095 R：通って

図 9 繰り返しの例

[1 : 小会議室 : 聞き返し]

o 0039 R : 部屋は小会議室は空いてますか

x 0040 L : {F と} 火曜日の {F え} 14 時半からは {F え} 小会議室は空いていません

図 10 質問－応答の例

図 9、図 10 の例の結果から分かるように、繰り返された発話と質問－応答など隣接ペアになっている箇所は、方式 1 では認識できない。そこで、方式 1 の結果から①、②の箇所を排除するように改良を行った（方式 2）。

方式 2

一人一人の発話の流れを見て、「はい」、「分かりました」などの実質的に意味をもたない単語をマークし、マークのない発話によってセグメントがはじまり、次にマークのない発話が現われるまでの間を区切る。これにによって示される区間の最初の発話が、前の発話に対して同じ表現（単語、句）の繰り返し、または質問－応答、依頼－受諾など隣接ペアでない場合、その区間をセグメントとする。

方式 2 の評価

方式 1 と同じく再現率と精度で方式 2 を評価する。

対話	再現率(%)	精度(%)	対話	再現率(%)	精度(%)
99atr01.ds1	100.0	78.3	99kyo0a.ds1	100.0	92.6
99atr02.ds1	100.0	95.7	99osa01.ds1	100.0	78.8
99chi0a.ds1	98.0	98.0	99osa02.ds1	98.5	100.0
99crl01.ds1	95.0	100.0	99osa0a.ds1	95.7	100.0
99crl02.ds1	96.6	93.3	99tsu0a.ds1	100.0	91.2
99crl03.ds1	97.1	100.0	99uec0a.ds1	95.8	85.2
99kyo01.ds1	100.0	100.0	99was03.ds1	100.0	95.1

表 3 方式 2 の再現率と精度

	再現率	精度
平均	98.3	93.4
標準偏差	1.94	7.7

表 4 方式 2 の平均と標準偏差

方式 2 に改良した結果、精度は 24.7%向上し、極めて高い数値を得ている。しかしな

がら、この方式でも境界を正しく認識できないところがあった。本方式で境界を正しく認識できないところには次のようなパターンが見られた。

- ・ セグメント境界の外にある発話に対する応答
- ・ 質問－（応答－了解）ⁿ（但し、**n** は 2 回以上）
- ・ 質問－応答－情報提供－受諾

図 11 で示した例で、発話 0021 は 0018 への応答だと考えられる。質問－応答 1－了解 1－応答 2－了解 2 のように、質問－応答のペアで複数の答えが求められるか、応答の内容がいくつかの段階に分けられるような質問に対して、その応答の間に了解が割り込まれるケースを質問－（応答－了解）ⁿ に分類した（図 12）。質問－応答－情報提供－受諾は、セグメントの中に 2 つの隣接ペアが存在するケースである。図 13 の例では、求められた情報を基に情報を提供（発話 0030）している。

以下に図 11～図 13 を示す。

[1：誕生日：割り込み]

x 0018 L: {F あの} {C 実は} ですね今日主人の誕生日なんです

[1：誕生日]

x 0019 L: {D それで} 主人の好きなものにメニューを合わせてやりたいなと思ひまして

0020 R: そうでございますか

o 0021 R: それはおめでとうございます

図 11 セグメント境界の外にある発話に対する応答の例

[2: 所要時間:]

o 0081 R: {F あの} 空港から {F え} {F その} 厚生年金会館まで {き[?] } どのくらい
かかりますでしょうか

x 0082 L: {F と} JR が 50 分ぐらい*地下鉄が 15 分ぐらい

0083 R: * {B はい}

0084 R: はい

x 0085 L: {D で} 徒歩が 10 分ぐらいです

0086 R: はい

図 12 質問－（応答－了解）ⁿ の例

[1: 講演会の日時: 聞き返し]

x 0028 L: 火曜日の {F え} 14 時から何時まででしょうか

o 0029 R: {F えと} 5 時半までで (す)

x 0030 L: {F えと} {こ[?]} {F え} じゅ水曜日からの出張のための {F え} 東京出張の
ための新幹線のお (や) くが {17[juushichi]} 時半から入ってますので {F と} その予定は登
録できません

0031 R: は

図 13 質問－応答－情報提供－受諾の例

3.3 考察

本研究で用いられた方式の結果を見ると再現率、精度ともに極めて高い数値が得られている。これほど高い数値が得られたことに対して、いくつか批判が予想される。それはまず、本研究で問題としている談話セグメンテーションというのは、もともと容易に解決できる程度のもものではなかったであろうかという批判である。もうひとつは、質問－応答の判断が本当に簡単にできるかという点である。

問題がもともと簡単だったのではないかという批判に答えるために、従来提案されている手がかり語による方式との比較を試みた。手がかり語として談話標識、フィラー、接続詞を考えてみることにする。本研究で利用されたデータには、手がかり語情報が付与されているので、それを利用して再現率と精度の計算を行った。まずは談話標識、フィラー、接続詞を、それぞれを単独基準としてセグメンテーションを行った。その後、これら手がかり語すべてを基準としてセグメンテーションを行い、再現率と精度を計算した。

対話	フィラー		談 話 標 識		接 続 詞		手がかり語全て	
	再現率	精度	再現率	精度	再現率	精度	再現率	精度
99atr01.ds1	22.2	36.4	16.7	100.0	5.6	33.3	44.4	47.1
99atr02.ds1	31.8	43.8	18.2	50.0	4.5	100.0	54.5	48.0
99chi0a.ds1	10.2	33.3	32.7	80.0	4.1	40.0	46.9	57.5
99crl01.ds1	15.0	12.5	5.0	100.0	10.0	40.0	30.0	20.0
99crl02.ds1	13.8	25.0	6.9	50.0	6.9	100.0	27.6	36.4
99crl03.ds1	22.9	50.0	5.7	100.0	5.7	66.7	34.3	57.1
99kyo01.ds1	68.8	34.4	12.5	50.0	0.0	0.0	81.3	36.1
99kyo0a.ds1	40.0	32.3	16.0	57.1	4.0	100.0	60.0	38.5
99osa01.ds1	73.1	33.3	11.5	37.5	3.8	100.0	88.5	33.3
99osa02.ds1	59.1	30.5	7.6	41.7	1.5	12.5	68.2	30.4
99osa0a.ds1	30.4	38.9	26.1	85.7	8.7	100.0	65.2	55.6
99tsu0a.ds1	3.2	6.3	6.5	50.0	0.0	0.0	9.7	15.0
99uec0a.ds1	50.0	24.5	8.3	40.0	0.0	0.0	58.3	25.0
99was03.ds1	10.5	33.3	18.4	87.5	0.0	0.0	34.2	65.0

表 5 手がかり語を利用した再現率と精度 (%)

	フィラー		談 話 標 識		接 続 詞		手がかり語全て	
	再現率	精度	再現率	精度	再現率	精度	再現率	精度
平均	32.2	31.0	13.7	66.4	3.9	49.5	50.2	40.4
標準偏差	22.7	11.4	8.2	24.3	3.3	43.6	22.0	15.2

表 6 手がかり語の平均と標準偏差

従来の研究（中里他，1999）で提案されていた手がかり語だけを頼りにセグメンテーションを行った結果、表 5 のような再現率と精度が得られた。この方法では再現率、精度ともになんて低い数値が結果として得られた。この比較で分かるように、対話のセグメンテーションはもともと簡単な問題ではない。

質問－応答などの判断は精度よくできるか、さらに計算機による自動化は可能かという疑問について説明する。方式 2 に改良するために分類したケースは質問－応答、依頼－受諾などの隣接ペアであった。質問－応答はさらにいくつかのパターンに分類される。典型的な質問－応答の他に、以下の 5 つのパターンがある。

- ・ 質問－応答（繰り返しを含む）
- ・ 質問－応答－繰り返し
- ・ 質問－応答／質問－応答
- ・ ステージング
- ・ 質問－応答*

質問－応答（繰り返しを含む）は、応答の中に質問の内容を繰り返すパターンで、質問－応答－繰り返しは、応答の内容を質問した者が繰り返すパターンである。質問－応答／質問－応答は、質問に対する質問が応答として発話され、それに対して応答するパターンである。ステージングは、住所、電話番号などをいくつかの桁に分けて話し、それを追っていく形式で応答が形成されるものである。質問－応答*は、応答の内容が質問の内容とは異なる提案をしたり、情報を提供する場合である。

[2：電車での行き方]

o 0010 R：{F え} 電車を使うほうだとどうなりますか

x 0011 L：電車を使う場合は {F え} JR と地下鉄を使う方法と* {服部[hattori]} 電鉄という私鉄と*地下鉄を使う方法があります

0012 R： * {B はい}

0013 R：* {B はい}

0014 L：はい

図 14 質問－応答（繰り返しを含む）の例

[1：数量]

x 0028 L：こちらのほう数量は

o 0029 R：は 1 点で

x 0030 L：1 点

0031 R：はい

図 15 質問－応答－繰り返しの例

[1：収容人数：聞き返し]

x 0005 L：{F と} どの会議室の収容人数でしょうか

o 0006 R：{F え} {F えと} いちおう全部お願いします

x 0007 L：{F と} 小会議室が 8 人中会議室が 5 人大会議室が 30 人セミナー室が 5 人第 1 演習室と第 2 演習室が 20 人講習室が 25 人となっています

0008 R：はい

図 16 質問－応答／質問－応答の例

[1：顧客番号]

x 0004 L：お客様番号をお願いします

o 0005 R：{F え} {1} {4} {6}

x 0006 L：{1} {4} {6}

o 0007 R：{5} {6} {5} *

x 0008 L：* {5} {6} {5}

o 0009 R：{2} {2} {3} です

x 0010 L：{2} {2} {3}

0011 R：はい

図 17 質問－応答（ステージング）の例

[2: ピラミッドの周囲での移動]

x 0100 L: に行ったら*ピラミッドである

0101 R: * {B うん}

o 0102 R: ああまずレモン {畑[batake]} がある

図 18 質問－応答*の例

このような構造を計算機によって 100%認識するのは難しいと考えられる。しかし、質問－応答でも繰り返しがあれば処理は可能であるだけでなく、質問－応答については、計算機で質問－応答を識別するアルゴリズムが提案されている（Kurohashi, 1994）。

提案した方式では、話者ごとに分けた発話の流れを見て実質的に意味の持たないと思われる発話がセグメントを決めるための基準であると判断する。一人一人の発話の流れを見てみると、対話では断片的で途切れていると思われたものが、ほぼ完全にいわゆる文を形成していることが多かった。この知見を基に考えられるいくつかの応用を提示する。

まずは、セグメンテーションの機械化に役立つと考えられる。本論文で提案した方式は、処理のアルゴリズムが極めて簡単明瞭であり、セグメント境界が精度よく取り出せる。また、計算機で処理できる文法を作成するためのデータとして活用できると思われる。これは一人ずつの発話を見ていくと、対話での発話に規則性が見つけやすいと考えられるからである。

第 4 章

セグメントの意味的側面

－主導権の概念から－

4.1 主導権の定義

3 章では、対話をセグメント化するための方式を提案した。しかしながら、そのセグメントがどのような意味をもっているものなのかについては議論してこなかった。本章では、セグメントの意味的な側面について主導権という観点から議論をする。

主導権の定義については、談話の流れと課題のどちらに焦点を合わせるかによって異なっている。本論文で紹介する主導権の定義は、(Robin Cohen et al., 1998) が

- ・ Whittaker and Stenton, 1988
- ・ Walker and Whittaker, 1990
- ・ Chu-Carrol and Brown, 1998
- ・ Smith, 1993
- ・ Guinn, 1996

の文献を基本として、整理・拡張したものである。

主導権を 4 つの定義にもとづいて説明する。

定義 1

この理論で説明される主導権の定義は対話の流れをコントロールする方に主導権

が与えられる。(Whittaker and Stenton, 1988) は、発話タイプを主張 (assertion)、命令 (command)、質問 (question)、促し (prompt) の 4 つのカテゴリに分類した。そして、各発話タイプに関する主導権の位置はコントロールルールによって決められる。

発話タイプに関するコントロールルールは、

- ・ 質問：相手の質問、命令に直接続く場合以外は、話し手が主導権を持つ
- ・ 主張：相手の質問への応答でない限り、話し手が主導権を持つ
- ・ 命令：命令調ではない間接的命令を含み、話し手が主導権を持つ
- ・ 促し：話し手がターンを放棄していると考えられるため、聞き手が主導権を持つ

と定義される。

主導権の移動については、話し手が主導権を相手に渡す際の言語的特徴として放棄 (abdication)、要約 (summary)、繰り返し (repetition) があると分析した。また、主導権を獲得する場合に割り込み (interruption) が使われると説明している。

割り込みが起きる状況については、情報の質とプランの質に関係する原理があると分析している。情報の質に関する割り込みは (a)、(b) の場合である。

(a) 聞き手は主張 P を信じ、P が適切だと信じている。そして、話し手が P でないものを信じるか P を知らないと、聞き手が信じる場合、割り込む

(b) 話し手の主張が適切ではあるが曖昧な場合、割り込む

プランの質に関する割り込みには (c)、(d) の場合がある。

(c) 聞き手は主張 P を信じている。そして、主張 P が提案されたプランに障害を引き起こすか、提案されたプランがすでに満たされたと信じる場合、割り込む

(d) 提案されたプランについての主張が曖昧な場合、割り込む

この原理に基づいて割り込みが行われると主張する。

図 19 の例を定義 1 で分析すると、A の質問 (発話 1、3) に対して B は応答する (発話 2、4) だけであるので主導権は常に A が持っている。

定義 2

現在解決しようとしている問題への解の候補（またはプラン）を提案している方が主導権を持つ。定義 1 と比較すると、定義 2 は対話の方向に関係するものではなく、課題に対して提案された解を構成しない発話は主導権を取らない。また、この定義では、談話の中に誰も主導権を持っていないケースを許す。

この理論における主導権の移動には以下の 3 つのパターンがある。

- ・問題を解くことができなくなった場合。主導権保持者にそれ以上は問題解決へとリードする能力がないときに起きる。
- ・相手が間違いに気づいて、修正を提案してくる場合。この場合、新しい案を提案することで主導権が移動する。
- ・相手が対案を出してくる場合。対案のメリットを基礎に主導権は移動する場合と移動しない場合がある。

図 19 の例を定義 2 に適用して主導権の分析を行う。発話 1 で A が問題を提示するが、まだ主導権は持っていない。B は発話 2 で解を提案することによって主導権を取る。B は、A が主導権移動の状態を示さないので続けて主導権を保持する。

-
1. A: CN タワーにはどうやって行ったらよいでしょうか。
 2. B: ガーディナー高速道路に着くまで 404 号を南へ行ってください。高速道路を西に行って、標識に従ってください。
 3. A: 高速道路はどのくらい走らないといけないのですか。
 4. B: 約 5~6km です。標識がちゃんとあるので問題ないはずです。
 5. A: 分かりました。ありがとう。
-

図 19 主導権分析のための例文 1 (Cohen et al., 1998)

定義 3

問題解決のドメインのゴールを提案する発話をするによって主導権をとると

する。この定義では、対話のコントロールと課題のコントロール両方を考慮する。定義 1 と異なる点は、ドメインのゴールを考慮するところである。また、課題に関する問題を解く発話であっても、それがゴールへ向かう対話の方向や新しいゴールを提案していない場合には主導権に結びつかないところが定義 2 との相違点である。

図 19 の例を定義 3 で説明する。発話 1 で、A がゴールを提案しているので主導権を取る。発話 2 で B が問題を解いているが、主導権を取ってはいない。B は発話 2 以後の対話をコントロールするが、単純に問題を解いていて、新しいゴールの提案がないため主導権を取ることはない。

定義 4

プロセス（process）と主導権の強度（strength）という新しい概念を導入した定義である。プロセスは対話の中の複数のターンのまとまりである。1 つのプロセスで、最初のターンをとることによって主導権を持つ。1 つのプロセスが他のプロセスと重複したり、割り込み（interruption）によって他のプロセスが割り込むケースもある。また、1 つのプロセスを終了せず他のプロセスがはじまる場合もある。対話は、参加者が貢献するこれらプロセスのまとまりである。

プロセスの最初のターンによって主導権を取るようになるが、すべてのターンには主導権の強さがあるとしている。ある閾値以上の強度をもつ発話は新しいプロセスを始められるが、ある閾値以下だと新しいプロセスを始めることはできないとされる。定義 1 で示されたコントロールルールによる主導権は典型的で強い主導権である。主導権の度合いが強い発話には以下のような発話がある。

- ・ 話題を変更することになる発話
- ・ 相手から要求されてではなく自発的に情報を提供する発話
- ・ 提案を求める発話（ただし、命令の発話よりは弱い）
- ・ 意志が強い、あるいは断定する発話

この定義を用いると、「主導権のある／なし」ではなく、全体に主導権があると見て

分析することができる。

図 19 の例を定義 4 に適用すると、図 19 は、発話 1～2 と発話 3～5 の 2 つのプロセスで構成される。各プロセスともに A によって始められるので主導権は A が取る。

ここで示された 4 つの定義を図 20 に示す対話の例に適用して、定義によって主導権がどう解釈されるかを説明する。

-
1. C: こちら回路修理の店です。どうでしょうか。
 2. U: 回路を直したいのです。
 3. C: 回路の番号は何番ですか。
 4. U: Rs111 です。
 5. C: その回路なら知っています。1 番と 7 番の LED が相互に光っているはずですが。
 6. U: LED は消えています。
 7. C: LED がついていないときスイッチはどうなっていますか。
 8. U: スイッチは入ってません。
 9. C: スイッチを入れてください。
 10. U: 分かりました。
 11. C: つまみを 1、0 にして下さい。
 12. U: 分かりました。
-

図 20 主導権分析のための例文 2 (Cohen et al., 1998)

定義 1 では、C がプランを提供することで主導権を取る。サブプランを出し、最後まで対話の流れを進めていくのでずっと主導権を取っている。U は単純な答えか促し (Okay) の発話しかしないので主導権を取ることはない。

定義 2 では、C が問題への解を提案する発話 5 で初めて主導権が起きる。C は問題解決の発話を続けて最後まで主導権を取る。

定義 3 では、発話 2 がゴールを提案する発話である。従って、最初は発話 2 で U が主導権を取ることになる。C は発話 3 で ID を使って新しいサブゴールを提案することで主導権を取る。C は発話 5 と 7 でもう一度主導権を取り、発話 8 以降は新しいサブゴールの提案がないので両方とも主導権が取れない。

定義 4 によると、例は 6 個のプロセスに落とすことができ、各プロセスの主導権を C が取っている。U の発話は自発的な情報提供がなく、C の質問／命令に答えるだけである。主導権の強さについては、C の発話は強くて、U の発話は弱い主導権しか持たないと解釈できる。

図 21 には、図 20 の対話例 2 に対して主導権の定義 1～4 に基づいて主導権を付与した結果を示す。

			定義番号:				
			1	2	3	4	
1.	C: こちら回路修理の店…	発話番号	1 C	—	—	『C	
2.	U: 回路を直したい…		2 C	—	U	』	
3.	C: 回路番号は…		3 C	—	C	『C	
4.	U: Rs111…		4 C	—	—	』	
5.	C: その回路なら…		5 C	C	C	『C	
6.	U: LEDは消えています…		6 C	C	—	』	
7.	C: LEDがついていない…		7 C	C	C	『C	
8.	U: スイッチは入って…		8 C	C	—	』	
9.	C: スイッチを入れて…		9 C	C	—	『C	
10.	U: 分かりました		10 C	C	—	』	
11.	C: つまみを1、0にして…		11 C	C	—	『C	
12.	U: 分かりました		12 C	C	—	』	

図 21 主導権の分析

この例では、定義 4 がプロセス（『』）という単位を持っていることを別にすれば定義 1 に似ている。定義 2 は主導権保持者を認めている点は似ているが、課題に焦点を当てて問題の解を提供することによって C が主導権を取っている点が違う。定義 3 はゴールと新しいサブゴールを提案しているところで主導権を取る。

4.2 主導権の観点からのセグメント

本論文で提案した方式では、主導権を持つ発話そのものではないが、主導権と関係しそうな発話を選んでいることになる。なぜなら、本方式では、基本的には実質的意味のない発話をセグメント境界だと見なすからである。主導権そのものではないとい

った意味は、例えば質問 - 応答の場合、主導権の定義によって、質問をした方の発話者に主導権があるとしたり、応答する発話者の方に主導権があるとしたり、または応答の内容によって主導権の保持者が変わることがあるからである。すなわち、主導権の定義はどこに焦点を置いているかによって結果が異なる。ここでは、主導権の 4 つの定義と得られたセグメントとの関係から、セグメントの意味的側面を観察する。

本論文で提案した方式で得られたセグメントは、それが正解であった場合、セグメントのはじめは意味のある発言であるため、主導権の定義 1~4 はすべてセグメントを始めた発話者に主導権があると判断する。但し、定義 2 と定義 3 は内容によっては誰も主導権を持たないので、主導権はある場合とない場合がある。また、繰り返しが含まれるセグメントは、対話の進展にはつながらないので、正解セグメントと同じ解釈をすることができる。

それ以外に現われるセグメントのパターンはどのように解釈されるだろうか。典型的なパターンとして質問-応答が考えられる。質問-応答の中でも両方とも意味のある発話をする場合、定義によって主導権保持者が大きく変わるので、例題を利用して分析することにした。図 22 で示したセグメントは、定義によって主導権を持つ話者が違ってくる実例である。話者 L と R は、両方とも内容のある発話をするので、方式 1 では正しくセグメント境界を認識することができなかったが、方式 2 によってセグメント境界が正しく認識された質問-応答である。

[1 : 行き先:]

- o 0028 R : {F あの} {F え} そこからどこ行きの切符を買えば良い良いんでしょうか
x 0029 L : {F え} そこからですと {F え} 一番近い方法ですと*寺町という所まで*快速が走っています
0030 R : * {B はい}
0031 R : * {B はい}
0032 R : はい
-

図 22 主導権分析のための実例

定義 1 で主導権を考えると、R が発話 0028 で質問をし、L が応答をするので、こ

こでの主導権は R が持つことになる。定義 1 は、発話の内容を考慮するより発話のタイプだけで主導権を定義しているからである。この例は、R が主導権を放棄する発話（はい）でセグメントが終わっている。

定義 2 は課題解決に依存している理論で、提案した解の性質のよっては誰も主導権を持たないケースが含まれるので、質問－応答の例では応答をする方に主導権があるか、両方とも主導権を持たないかの 2 つのケースが考えられる。図 22 の例は、提示された問題について L が解の提案をするので、L が発話 0029 で主導権を持つと解釈される。

定義 3 はゴールに依存し、問題解決のためのサブゴールを提示する方に主導権があるとした理論である。従って、質問をしてゴールを提供するほうか、ゴールへの解としてサブゴールを提示する応答をする方に主導権があると考えられる。また、質問と応答が上記の両方である場合はセグメントの中で両方が主導権を持つことも考えられる。この例では、R は発話 0028 でゴールを提示することにより主導権を取っていると判断される。次の発話で L は問題を解いている。単純に問題を解くだけではなく「快速」というサブゴールを提示することにより発話 0029 では L が主導権を取っている。

定義 4 はプロセスを始める方に主導権があるとしながらも、すべての発話には主導権があり、主導権の強度が異なるとしている。セグメントはプロセスに当たると考えられる。従って、この例では R がプロセスの最初の発話をするので主導権を持つことになる。しかし、L の発話が定義 4 で示した「強い主導権を持つ発話」であるため、このセグメントは両方とも強い主導権を持つと考えられる。

以上では、対話のセグメントに定義 1～4 までを適用して主導権を分析した。主導権の定義が 1 つに定まっていない現状の中では、この概念から分析するセグメントの意味も定義の数だけできるのかも知れない。どのような目的をもって開発するかによって主導権の定義を使い分けるか、または直感に合う主導権の定義をする必要があると考えられる。いずれにせよ、今回のようなセグメントに区切っておくことにより、主導権を考えると対話の全部を見る必要がなく、問題になるような箇所だけ（質問－応答のような）を分析して、自分に合うものを選択することができるのである。セグメントの意味的側面を主導権の概念を利用して考察して、本論文で得られたセグメントは主導権を考えるための基礎データを作るために役立つと考えられる。

第 5 章

おわりに

本論文では、対話分析の単位となるセグメントを精度よく取り出す方法を提案し、この方式でのセグメンテーションが従来の研究で利用された手がかり語によるセグメンテーション方式より精度よくセグメンテーションができることを示した。

分析に利用した対話データは、課題のバリエーションのある 14 対話である。対話の発話を一人ずつの発話に分け、その発話の流れの中で実質的には意味を持たないと判断される「はい」などの発話で対話が区切られると考え、その考え方によってセグメントの境界を取り出すのが方式 1 である。すなわち、意味を持つ発話間が 1 つのセグメントを形成するという考え方である。対話データに付与されているセグメントを正解と考えて、本論文で提案した方式によって見つけたセグメントと比較した。方式 1 の結果で得られた再現率は平均 98.3%、精度は平均 68.7%と、再現率は高いが精度は予想より低いことが分かった。誤ってセグメント境界を判断するところを分析して、特徴として、繰り返しと質問－応答などの隣接ペアの 2 つのパターンが多く見られることを明らかにした。分析の結果に基づいて方式の改良を行い、方式 2 を提案した。方式 2 に改良した結果、精度が 24.7%向上し、93.4%の極めて高い数値を得ることができた。

談話セグメントの意味的側面を考えるために、主導権という観点からセグメントを分析した。具体的には、セグメントに区切られた発話と主導権を、4 つの定義によってどのように解釈されるかを考察した。主導権の定義は観点によってことなるが、それぞれの目的に合った定義を分けて使うことになる場合でも、問題になるようなところだけを分析することを可能にするという意味で作業の効率を高めることができると考えられる。

謝辞

本研究を進めるにあたり、石崎雅人助教授には、数々のご指導・ご助言をいただきました。また、情報科学研究科自然言語処理学講座の望月源助手からは、セグメント方式を考える上での有用なヒントをいただきました。

ここに謝意を表します。

参 考 文 献

- (荒木他, 1999) 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人, 発話単位タグ標準化案の作成, 人工知能学会誌, Vol.14, No.2, pp.251-260, 1999.
- (Austin, 1962) John L. Austin, How to Do Things with Words, 1962
- (Cohen et al., 1998) Robin Cohen, Coralee, A., Christian, C., Mark, F., Kinson, H., Bowen, H., Celine, L., Fletcher, L., Nancy, M., David, P., Alex, Q., Saheem, S., What is Initiative? , User Modeling and User-Adapted Interaction 8, pp.171-214, 1998.
- (コーパス利用 WG, 2000) 人工知能学会 談話・対話研究におけるコーパス利用研究グループ, 様々な応用研究に向けた談話タグ付き音声対話コーパス, 人工知能学会 研究会資料 SIG-SLUD-9903, 2000.
- (Coulthard, 1985) Malcolm Coulthard, An Introduction to Discourse Analysis, second edition, Longman, London, 1985.
- (Grosz and Sidner, 1986) Babara J. Grosz and Candace L. Sidner, Attention, Intentions, and the Structure of Discourse, Computational Linguistics, Vol.12, pp.175-204, 1986.
- (Halliday, 1976) M. A. K. Halliday and Ruqaiya Hasan, Cohesion in English, Longman, London, 1976.
- (Kurohashi, 1994) Sadao Kurohashi and Makoto Nagao (1994), Automatic Detection of Discourse Structure by Checking Surface Information in Sentences, Proceedings of the 15th International Conference on Computational Linguistics, pp.1123-1127, 1994.
- (Levinson, 1983) Stephen C. Levinson, Pragmatics, Cambridge University Press, Cambridge, 1983.

- (Nakatani et al., 1995) Cristine H. Nakatani, Babara J. Grosz, and Julia Hirschberg, Instruction for annotating discourse, Technical report, Center for Research in Computing Technology, Harvard University, 1995.
- (中里他, 1999) 中里収, 田本真詞, 菊池英明, 吉村隆, 課題遂行対話における対話潤滑語の認定, 人工知能学会誌, Vol.14, No.5, pp.140-146, 1999.
- (Schegloff and Sacks, 1973) Emanuel A. Schegloff and Harvey Sacks, Opening up Closings, *Semiotica*, Vol.8, pp.289-327, 1973.
- (Searle, 1969) John R. Searle, *Speech Acts: an Essay in the Philosophy of Language*, Cambridge University Press, 1969.
- (Searle, 1976) John R. Searle, *Expression and meaning*, Cambridge University Press, 1976.
- (Stenström, 1994) Anna-Brita Stenström, *An Introduction to Spoken Interaction*, Longman, London, 1994.
- (Walker and Whittaker, 1990) Walker, M. and S. Whittaker, Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation, In the Proceedings of Association for Computational Linguistics, 1990.
- (Whittaker and Stenton, 1988) Whittaker, S. and P. Stenton, Cues and control in expert-client dialogue, ACL88, In the Proceedings of Association for Computational Linguistics, 1988.
- (山下他, 1999) 山下洋一, 小磯花絵, 堀内靖雄, 音声対話に対する談話セグメントのタグ方式の検討, 人工知能学会誌, Vol.14, No.2, pp.282-289, 1999.