JAIST Repository
https://dspace.jaist.ac.jp/

Title	対話セグメント分割に関する研究
Author(s)	小倉，加奈代
Citation	
Issue Date	2001-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/335
Rights	
Description	Supervisor:石崎 雅人，知識科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修　士　論　文

対話セグメント分割に関する研究

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

小倉　加奈代

2001 年 9 月

Copyright _ 2001 by Kanayo Ogura

目　次

1. はじめに
1.1 研究の背景----------1
1.2 研究の目的----------2
1.3 本論文の構成----------2

2. 関連研究
2.1 セグメンテーション----------4
2.2 手がかり語を用いた手法----------6
2.3 実質的に意味のない単語を用いた手法----------8
2.4 テキストタイリングアルゴリズムを用いた手法----------10
2.5 重み付けを用いた手法----------12

3. セグメンテーション手法の評価
3.1 使用データ----------14
3.2 実験の評価方法----------18
3.3 実験の方法----------20

4. 手がかり語を用いた手法についての評価、考察
4.1 手がかり語を用いた手法
4.1.1 機能を重視した手がかり語を用いた手法----------22
4.1.2 形態素解析の結果の手がかり語を用いた手法----------24
4.2 手がかり語を用いたセグメンテーションの方法----------27

4.3 手がかり語を用いた手法の結果と考察----------27

5. 実質的に意味のない語を用いた手法と評価、考察
5.1 実質的に意味のない語を用いた手法----------33
5.2 実質的に意味のない語を用いたセグメンテーションの方法----------33
5.3 実質的に意味のない語を用いた手法の結果と考察----------35

6. テキストタイリングアルゴリズムの手法と評価、考察
6.1 テキストタイリングアルゴリズムを用いた手法----------41
6.2 テキストタイリングアルゴリズムを用いたセグメンテーションの方法--------44
6.3 テキストタイリングアルゴリズムを用いた手法の結果と考察----------45

7. 重み付けを用いた手法と評価、考察
7.1 重み付けを用いた手法----------57
7.2 最大エントロピーを用いたセグメンテーションの方法----------57
7.3 最大エントロピーを用いた手法の結果と考察----------59

8. おわりに----------65
謝辞----------67
参考文献

図　目　次

図 2.3　実質的に意味のない発話リスト（一部）----------8
図 3.1.1　対話潤滑語タグが付与されているデータ例----------15
図 3.1.2　品詞情報が付与されているデータ例----------16
図 3.1.3　談話行為タグが付与されているデータ例----------16
図 3.2.1　再現率の算出方法----------18
図 3.2.2　精度の算出方法----------18
図 3.2.3　それぞれのセグメントの関係図----------19
図 3.3　すべての手法で共通するアルゴリズム----------20
図 4.1　接続標識、談話標識、フィラーを用いた手法の例----------23
図 4.1.2　接続詞、感動詞、フィラーを用いた手法の例----------25
図 5.1　実質的に意味のない語を用いた手法の例----------34
図 5.3　36 対話でマッチした実質的に意味のない語リスト----------40
図 6.1.1　ブロックの設定、および隣接するブロックの設定について例----------42
図 6.1.2　類似度の計算式----------43
図 6.1.3　セグメント決定の条件式----------43
図 6.1.4　新出単語の頻度を用いてセグメンテーションを行なう手法の類似度---44
図 7.2.1　最大エントロピーの素性----------58
図 7.2.2　セグメンテーションの決定規則----------59

表　目　次

表 4.3.1　接続標識、談話標識、フィラーを用いた結果----------27
表 4.3.2　接続詞、感動詞、フィラーを用いたセグメンテーションの結果 1--------28
表 4.3.3　接続詞、感動詞、フィラーを用いたセグメンテーションの結果 2--------29
表 5.3.1　実質的に意味のない語を用いたセグメンテーションの結果 1----------36
表 5.3.2　実質的に意味のない語を用いたセグメントの結果 2----------37
表 5.3.3　実質的に意味のない語を用いたセグメントの結果 3----------38
表 6.3.1　テキストタイリングアルゴリズムでのセグメンテーションの結果 1------45
表 6.3.2　テキストタイリングアルゴリズムでのセグメンテーションの結果 2------46
表 6.3.3　テキストタイリングアルゴリズムでのセグメンテーションの結果 3------48
表 6.3.4　テキストタイリングアルゴリズムでのセグメンテーションの結果 4------49
表 6.3.5　テキストタイリングアルゴリズムでのセグメンテーションの結果 5------50
表 6.3.6　テキストタイリングアルゴリズムでのセグメンテーションの結果 6------51
表 6.3.7　テキストタイリングアルゴリズムでのセグメンテーションの結果 7------53
表 6.3.8　テキストタイリングアルゴリズムでのセグメンテーションの結果 8------54
表 7.3.1 最大エントロピーを用いたセグメンテーションの結果 1----------59
表 7.3.1 最大エントロピーを用いたセグメンテーションの結果 2----------61
表 7.3.1 最大エントロピーを用いたセグメンテーションの結果 3----------62

第１章
はじめに

1.1研究の背景

近年、インターネットや携帯電話の爆発的な普及により、メールやチャットといった新しいコミュニケーション形態が汎用的に使われつつある環境になっている。現状としては、若年層でメールやチャットの普及率が高いが、今後は、若年層以外の年代においても普及する可能性が高いと推測できる。それに従い、メールやチャットといった新しいコミュニケーション形態は、日常だけではなく、教育環境、企業内の情報インフラとして利用されることも推測される。そういったことを考え合わせると、新しいコミュニケーション形態での情報をうまく活用する技術を確立することが求められる。

また、こういった現状を受けて、人間対機械とのコミュニケーションの研究が多く見受けられるようになっている。人間対機械のコミュニケーションは、我々の日常のコミュニケーションと比較すると、スムーズにやりとりが成立するとは言い難く、正確にコミュニケーションが成立しているとは言えないのが現状であり、現在も多くの研究者たちが、より自然に人間対機械のコミュニケーションを行なうことのできる研究に取り組んでいる。しかし、これらの研究は、我々が対話を行なっている場合に無意識ではあるが、少し意識すると行なっていると認識できるような要素を利用しておらず、我々が普段交わすような対話を対象とはしていないものが多いことから、自然なやりとりということを念頭においた場合、より日常に近い対話を扱い、我々が少し意識してみると行なっていると認識できるようなレベルの要素を利用するという方向性が考えられる。結局のところ、対話に見られる人間対人間

のコミュニケーションは、人間対機械とのコミュニケーションと同一ではないにしろ、そのメカニズムを解明することにより、人間対機械とのコミュニケーションを改善し、新しい形態のコミュニケーションに対応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られる。

1.2研究の目的

本研究の目的は、人間対人間のコミュニケーションの原型ともいえる対話から、我々が少し意識すると行なっていると認識できるような要素を用いて、セグメントとよばれる話題のまとまりをコンピュータを用いて自動抽出をより精度よく行なう技術を確立する知見を得ることである。

本論文は、話題のまとまりを抽出する際に利用する要素として、主に、品詞情報等の表層的手がかりと、対話中に利用される単語の分布に着目した内容的手がかりの２つの要素を用いた手法をそれぞれ評価し、その最適な組み合わせについて実験的に明らかにする。

1.3本論文の構成

第 2 章では、本研究と関連する研究をまとめ、本研究で具体的にどういった方法を採用するかについて明確にする。

第 3 章では、本論文で利用するデータについて説明するとともに評価方法について述べる。

第 4 章では、手がかり語を用いたセグメンテーションについて手法と評価、考察を行なう。

第 5 章では、[金 00]をもとに、実質的に意味のない語を用いたセグメンテーションの手法を提案し、評価、考察を行なう。

第 6 章では、テキストタイリングの手法について評価、考察を行なう。

第 7 章では、重み付けを用いた手法について評価、考察を行なう。

第 8 章では、本研究のまとめを述べる。

第 2 章
関連研究

本章では、最初にセグメンテーションとは何かについて説明し、セグメンテーションを行なう際の手法に関して、手がかり語を用いた手法、実質的に意味のない単語を用いた手法、テキストタイリングアルゴリズムを用いた手法、重み付けを用いた手法の 4 つに関する関連研究を紹介する。

2.1セグメンテーション

セグメントとは、話題境界、談話境界と呼ばれるもので、話題のまとまりをなすもので、対話の分析の単位ともなりうるものである。

セグメンテーションの利点は、話題のまとまりを正しく見つけることにより、例えば情報検索では、より細かいレベルで関連する情報を抽出できたり、情報要約ができたりするという点があげられる。文章や独話の場合に、ある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ており、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[Grosz 86]。また、ある程度のまとまりごとに理解するという方法は我々が直感的に日常行なっている副目標→目標というような、小さなまとまりをもう少し大きな単位として扱うという行為に見られることである。

セグメントとして、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)、社会学の領域で扱われている会話分析の隣接ペア(adjacency pair)、言語学の領域で扱われている談話分析の交換構造(exchange structure)がある。これらについて、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。隣接ペアおよび談話構造については、それらが発語内行為または談話行為に基礎を

置いているが、これらの分類を高い精度で行なうのが難しいといった問題点がそれぞれにある。

仮に、精度よく、かつ、簡単にセグメンテーションを行なえた場合、計算機による様々な処理がしやすくなることが期待できる。例えば、対話分析のためのデータ作成や、知識ベースを用いた質問―応答システムなどのシステム作成のデータとして役に立つことが予想される。

2.2手がかり語を用いた手法

表層的な手がかりとして手がかり語の 1 つである談話標識を自動抽出し、それをもとにセグメンテーションを行なう手法がある[中里 99]。

この研究は、対話データの共有化を目指すための１つの試みとしてなされたもので、まず、相槌（「はい」、「うん」など）、フィラー（場をつなぎ、発話権の維持をする機能をもつ語。「ええと」、「あの」など）、談話標識（問題解決には直接関与せず、話題の転換や再開などを表わす機能をもつ語。「ところで」、「まず」、「つぎに」など）で構成される対話潤滑語と呼ばれるもののタグ付けに関する実験を行ない、次に、タグそのものの応用例として、対話潤滑語の中の談話標識を用いてセグメンテーションの実験を行なっている

まず、対話潤滑語のタグ付け実験においては、対話潤滑語としてどのような語が認定されるかが作業者間でどの程度一致するかに関する実験と、その一致において高い一致を示す語彙を手がかりとして対話潤滑語を半自動抽出することの 2 つを行なっている。

1 つ目の実験では、対話潤滑語のそれぞれに定義を与え、それをもとに作業者が対話データの各発話に対して対話潤滑語とするかしないかの判断、認定を行なっている。ただしここでは、フィラーを談話標識に含み、相槌と談話標識の 2 つのみを判断、認定の対象としている。また、作業者は 12 人で 1 対話あたり 4 人の割り当てで全 6 対話において行なっている。この実験の結果、全体として相槌、談話標識ともに、対一致率は平均 0.4 前後と低い数値をしめしているが、語彙ごとの一致率を見た場合、相槌の上位 5 種（「は」、「ない」、「と」、「え」、「はい」）と談話標識の 12 種（「いや」、「うん」、「えとじゃ」、「と」、「え」、「じゃ」、「あの」、「あじゃ」、「ですから」、「では」、「えと」、「あ」）が語彙全体の一致率を上回り、4 人中 3 人以上の一致が見られるという結果を示した。これにより、対話潤滑語の中には比較的多くの作業者間で一致するものとそうではないことがわかった。

2 つ目の対話潤滑語の半自動抽出では、1 つ目の実験において語彙全体の一致率を上回り、語彙全体の 1%以上の出現率のあったものと、データの頻度は少なかったが、相槌、談話標識の機能が明確であると考えられるものと、さらに、「それでは」の「で

は」や「それじゃ」の「じゃ」などのように「それ○○」の「それ」が省略された形のものを候補とした。それらによって、対話潤滑語として使われる可能性のある語をあらかじめ登録、限定することで、テキストからほぼ自動的に対話潤滑語が抽出できるようになり、作業者間の揺れも少なくなるというメリットを得ることができた。

次に対話潤滑語のタグ付けの応用として、１つ目に前の実験で用いた対話とは別の対話 5 対話を用いて 1 対話 3 名の割り当てで談話標識のみの自動抽出を行ない、2 つ目に自動抽出した談話標識の前・後ろ・それ以外の箇所でのセグメントの有無を調べるという２つを行なった。

1 つ目の談話標識の自動抽出では、5 対話中 376 発話のうち、9 種 46 語の自動抽出に成功した。

2 つ目の自動抽出した談話標識の前・後ろ・それ以外の箇所でのセグメントの有無の調査については、談話標識の前では 64%がセグメントであり、談話標識の後ろでは、88%の割合で同じ話題が続いており、談話標識が発せられた直後は話題が変わることがむしろ少ないということがわかった。

この研究で、セグメントに関して言えば、語彙情報をもとに自動抽出した談話標識語とセグメントには高い相関が見られ、自動抽出した談話標識語を用いて談話構造の自動推定が可能であるということが結論づけられている。

また、手がかり語には談話標識の他、接続標識、フィラー等も含まれており[人工知能学会談話対話におけるコーパス利用研究グループ 99]、これらを表層的な手がかりとしてセグメンテーションを行なうことも可能であると考えられる。

2.3実質的に意味のない語を用いた手法

セグメントを考える際に、対話の流れを見てみると、一見断片的に見える。しかし、一人一人の発話の流れをみた場合、発話の流れは断片ではなく、ある程度のまとまりをなしており、そのまとまりは「はい」などといった相手の応答をしめす発話、実質的に意味のない語で区切られている。そこで「はい」などの実質的に意味のない語を使ってセグメンテーションを行なう手法が提案されている[金 00]。

この研究では、本研究でも使用する人工知能学会などが提供する 14 対話の対話データの話者ごとに別々にファイルされているデータを用いて、話者それぞれの発話の流れを見て、直感的に実質的に意味のないと考えられる発話とそうでない発話を区別する作業を行なっている。そしてその作業から抜き出した実質的に意味のない発話以外の発話をセグメントとするという方式 1 を設定することにより、再現率 98.3%、精度 68.7%でセグメンテーションを行なうことができた。なお実質的に意味のない発話と認定された語のリストの一部を図 2.3 に示す。

「ありがとうございます」、「はい」、「ああそうですか」、「あはい」、「はいかしこまりました」、「はい分かりました」、「いえ」、「ありがとうございました」、「そうですか」、「なるほど」、「ええ」、「あそうですか」、「あそうでございますね」、
「はあなるほど」、「そうでございますか」、「どうもありがとうございます」、「はい充分です」、「いえとんでもございません」、「とんでもありません」、「うん」、「うんうんあるある下に」、「はいはいはい」、「え」、「あうんうん」、「はいはいはいはい」、「うん分かった」、「ないないない」、「ないか」、「じゃそれ」、「そう」、「ふん」、「あるあるあるある」、「うんと」、「うんない」、「うんうんうんうんうんうん」、「ない」、「あるあるあるあるうん」、「あるあるある」、「あそれはないこっちにはうん」、「あん」、「たぶんこう」、「うんとお」、「うんうんうんうん」、「うんはい」、「いや」、「だって」、「えあああそうか」、「うんうんうん」、「とし」、「そうそうそう」、「ね」、「うんふんふんうん」、「そうだよね」、「そうね」

図 2.3　実質的に意味のない発話のリスト（一部）

また、実質的に意味のない発話以外の発話をセグメントとするという規則において、再現率は高いが、精度が低いという結果が得られたことから、誤ってセグメントの判断を行なっている箇所の分析を行なった結果、同じ表現（単語、句）の繰り返し、質問－応答、依頼－受諾、開始と終了の挨拶などの隣接ペアのある箇所を方式 1 では認識できないことがわかった。そこで、方式１において、セグメントと判断した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、開始と終了の挨拶などの隣接ペアでない場合をセグメントとするという方式 2 を設定することにより再現率 98.3%、精度 93.4%でセグメンテーションを行なうことができた。方式 1 から方式 2 へ改良した結果精度は 24.7%向上し、極めて高い数字を得ている。しかし、この方式でもセグメントを正しく認識できない箇所を分析した結果、セグメント境界外にある発話に対する応答、質問－（応答－了解）（ただし（応答－了解）は 2 回以上の繰り返し）、質問－応答、情報提供－受諾のパターンであることがわかった。

この研究で提案された方式では、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され、この知見を基に、セグメンテーションの機械化や、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の活用に応用できると考えられている。

2.4テキストタイリングアルゴリズムを用いた手法

手がかり語、および実質的に意味のない語を用いた手法では、表層的な手がかりのみを用いている。しかし、話題の変化により、対話中で発せられる単語が変化するといったことは予測されることであり、こういった内容的な手がかりを用いる手法が有効である可能性がある。内容的な手がかりを用いる手法として、単語の頻度や分布を用いることで、話題の変化する位置を決める、つまりセグメンテーションを行なうテキストタイリングアルゴリズムがある[Hearst 94][Hearst 97]。

テキストタイリングアルゴリズムは、ある話題から別の話題へ推移する単語が文書中のどの辺りにあるかを探すということがコンセプトとされていて、まさに、そのどの辺りにあるかを探すということがセグメンテーションを行なうということである。

このアルゴリズムは、情報検索に主に利用されており、文書を対象に利用されているものであるが、文章内の一文の長さはそれぞれの文によってかなり異なっているため、予め、文書をトークン列と呼ばれる単語列単位に分割してブロックを設定する。この際、[Hearst 97]では、1 トークン列を 20 語と設定している。そして、1 トークン 20 語と設定されたブロック間の単語の索引語頻度、次のブロックに移行する際の新出単語の頻度から類似度を算出して、セグメンテーションを行なうものである。

このアルゴリズムは、cohesion scorer、depth scorer、boundary selector の 3 つの要素から主になりたっている。

Cohesion scorer は、話題の連続量、ブロックとブロックの間の類似度、つまり、前のブロックと後のブロックで話題が続いているのかどうかがどれくらい確かかを測る基準である。この cohesion scorer を決める際に主に 2 つの方法がある。1 つは、Block Comparison と呼ばれるもので、これは、隣り合うブロックから算出した索引語頻度を用いる方法である。もう 1 つは、Vocabulary Introduction と呼ばれるもので、隣り

合うブロック内の新出単語の頻度を用いる方法である。主に 2 つの方法が存在するが、このうち Block Comparison のほうが、有効であると示されている[Hearst 97]。

Depth scorer は、対象とするブロックとブロックの間の類似度を周辺の類似度と比較して決定する。

Boundary selector では、そこまでに算出したそれぞれの類似度から平均と標準偏差を算出し、平均－（定数×標準偏差）よりも値の大きい類似度をもった境界をセグメントとして選択する。

[Hearst 1997]では、このアルゴリズムで選択したセグメントと、人間が直感で判断したセグメントとの間には共通する部分が多くみられると評価を下している。

2.5 重み付けを用いた手法

本章の前節まで、表層的な手がかり、内容的な手がかりを用いたセグメンテーションの方式のうち、いくつかを上げてきた。そこで、次に、いくつかの方式を組み合わせてセグメンテーションを行なうための知見として、複数の表層的手がかりを用いた文書を対象としたセグメンテーションに関する研究[望月 99]を上げることにする。

この研究では、それぞれの表層的手がかりの重み付けを訓練データを用いた統計的手法により自動的に行なう手法と、複数の表層的手がかりの中で、実際にセグメンテーションを行なう際に有効な手がかりだけを選択することで訓練データへの過適合を避ける手法を提案している。

まず、表層的手がかりとしては、主語を表わす助詞、接続詞、照応表現、主語の省略、同一タイプの文の連続、語彙的連鎖、語彙的連鎖内の単語につく修飾語の変化を使用し、それぞれの手がかりに、各文間のセグメント境界への成り易さもしくは成り難さを示す文間のスコアを計算するためのスコアを与える。

次にそれぞれの手がかりのスコアに重要度に応じた重み付けを自動的に行なうために、正解セグメントの情報が付与されたテキストを訓練テキストとして、重回帰分析を使用して、各手がかりの重みの推定を行なう。

さらに、訓練データの量に比べて、表層的手がかりが多過ぎる場合に発生する過適合を解消するために、パラメータ選択手法の 1 つであるステップワイズ法を用いて良い推定ができると判断されたパラメータを加え、逆に別のパラメータが加えられたことにより、良い推定に役立たなくなったと判断されたパラメータを除去するという処理を繰り返し、最終的に有効なパラメータの組を選択する。

最後にこれらの評価実験として、日本語の国語の問題週から意味の切れ目を問う問題に使用された 14 テキストを用いて、精度、再現率を算出している。この実験では、語彙的連鎖以外の手がかりによる実験（実験 1)、語彙的連鎖のみの手がかりによる実験（実験 2)、設定した全ての手がかりによる実験（実験 3)、訓練テキストに

より自動的に計算された重みを使用し、人手によって重みを与えた実験（実験 4)、選択された手がかりのみを使用して自動的に重みを決定した場合の実験（実験 5）の 5 つの比較実験を行なっている。結果としては、実験 1～実験 3 を比較した場合、実験 3 が最も良い精度を引き出し、複数の手がかり（主語を表わす助詞、接続詞、照応表現、主語の省略、同一タイプの文の連続、語彙的連鎖、語彙的連鎖内の単語につく修飾語の変化）を組み合わせることが有効であることがわかっている。また、実験 3 を人手によって重みを与えた実験と実験 4 を比較した場合、実験 4 のほうが概ね良い精度を引き出し、自動的に学習された重み付けは、人手による手間を省き、客観的な値を得ることができることも含め、人手による重み付けよりも良い手法であるといっている。さらに、実験 4 に対して全ての手がかりを使用した実験と実験 5 を比較した場合、実験 5 が最もよい精度を引き出し、この研究で使用したパラメータ選択手法によって訓練テキストへの重みの過適合の問題が解消されていることを示している。

第 3 章
セグメンテーション手法の評価

本章では、実験に使用するデータについて説明するとともに、評価方法、実験方法について述べる。

3.1 使用データ

使用データは、大きく 2 つの課題遂行対話を使用する。

1 つは、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、父親当てクイズなどの課題を扱った全 14 対話であり、うち、対話潤滑語タグ、品詞タグ、談話行為タグのタグづけしてある 3 種類の対話データが存在する。

もう 1 つは、会議室に関するスケジューリングを扱った全 36 対話である。この対話に関しては、すべて 2 人対話である。これらに関しては、品詞タグ、談話行為タグのタグづけをしてある 2 種類の対話データが存在する。

それぞれのタグ付けされているデータに関してであるが、まず、対話潤滑語タグの付与されているデータを下に示す。

[2:登録依頼:]

00:00:180-00:02:540 0000 L:{F あ}{D それでは}予定の登録をお願いします

00:03:160-00:03:430 0001 R:はい

[2:教室会議:]

00:04:590-00:11:320 0002 R:{F えと}教室会議の*日時と場所を設定したいんですけど(も)

図 3.1.1　対話潤滑語タグが付与されているデータ例

この形式のデータには正解セグメントの情報と、発話の開始、終了時間、発話番号、発話者を区別するための記号、対話潤滑語タグが含まれている。

正解セグメントに関しては、すべての種類のデータに付与されているが、正解セグメントとするものには 1 または 2 の数字がふられている。この 1 と 2 の違いであるが、これは、1 の場合は関連性が強い場合で、2 の場合は、内容の変化が大きい場合とされている[山下 99]。

また、このデータでのセグメントでは、上で述べた内容の変化度合い、話題境界らしさ度合いのほかに、話題名、セグメント関係情報で構成されている。このセグメント関係情報には、「聞き返し」、「割り込み」、「復帰」がある。

対話潤滑語タグに関しては、この種類のデータのみに付与されている情報であり、図 8 の:{F あ}、{D それでは}のようなものである。対話潤滑語は前で説明したように接続標識、談話標識、フィラーの 3 つからなるものであるが、ここでは、C が接続標識に対応し、D が談話標識に対応し、F がフィラーに対応するようになっている。

次に品詞情報が付与されている対話データに関してであるが、これも下にデータの一部を示す。

---

[2] x

0 0 L あの フィラー

0 1 L すみません 感動詞

EOS

[1] x

1 2 L 女将 名詞-一般

1 3 L さん 名詞-接尾-一般

1 4 L いらっしゃる 動詞-自立

1 5 L ます 助動詞

1 6 L か 助詞-終助詞

EOS

図 3.1.2　品詞情報が付与されているデータ例

この形式のデータには正解セグメントの情報が括弧でくくられていて、前出の実質的に意味のない単語の認定のために使われた o、x の記号が含まれている。そして、左側に発話番号、その隣りに全データの通し番号、話者を区別するための記号、発話内容、形態素解析の結果の順で記されている。

最後に談話行為タグが付与してあるデータに関してであるが、これも下にデータ例を示す。

% C ----------

% L (働掛) 未知情報要求 (応答) - (セグメント) 2

% L

% C ----------
3 S: <1>なにか、その、使いたい会議の、条件を、会議とその条件を言ってやってください。

% C ----------
% L (働掛) - (応答) - (セグメント) -
% L
% C ----------
4 E: えーっと、まずですねー、演劇部としておこないたい会議はー、えー、演劇部全体の打ち合わせを３時間ほどおこないたいと、

図 3.1.3　談話行為タグが付与されているデータ例

この形式のデータには、L で始まる行に、談話行為の種類、正解セグメントの情報が記述されており、数字から始まる行には、発話番号、話者区別のための記号、発話内容が記されている。

本研究では、談話行為の種類を利用しなかったが、この談話行為について説明すると、談話行為は、相手に応答を求める発話の分類である働きかけ、相手の働きかけに応える発話の分類である応答に分けて記述されている。

本研究で利用した対話データ数は、本節で説明してきた様々な課題を含んでいる 14 対話と会議のスケジューリング課題を扱っている 36 対話の計 50 対話である。

3.2 実験の評価方法

本研究のすべての実験は、正解セグメントと、それぞれの手法においてプログラムを実行した結果取り出したセグメントの候補との再現率と精度を算出することにより評価を行なう。よって評価に必要な再現率と精度に関する説明を先に行なうことにする。

本研究のすべての実験では、対話データに付与されているセグメント情報を正解と考えて再現率、精度で評価を行なう。

まず、再現率であるが、これは、正解セグメント数（A）の中にどれくらい正解セグメントの中から提案する手法で候補として抜き出してきたセグメント数（B）があるかをパーセンテージで算出したものである。これを式にすると以下のようになる。

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

図 3.2.1　再現率の算出方法

次に精度であるが、これは、提案する手法で候補として抜き出してきたセグメント数（C）における、正解セグメントの中から提案する手法で候補として抜き出してきたセグメント数（B）の割合である。

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

図 3.2.2　精度の算出方法

また、よりわかりやすくするために、正解セグメント数（A)、正解セグメントの

中から提案する手法で候補として抜き出してきたセグメント数（B)、提案する手法で候補として抜き出してきたセグメント数（C）を図に表わすと下のようになる。

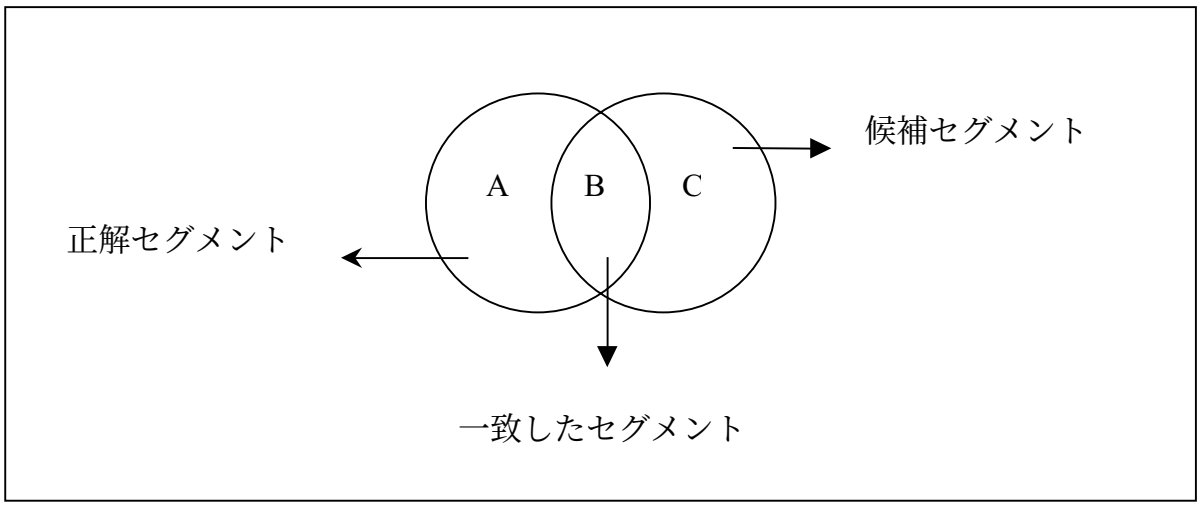

図 3.2.3　それぞれのセグメントの関係図

3.3 実験の方法

本研究のすべての評価実験は、再現率と精度を算出することにより評価を行なう。よってすべての手法で共通するアルゴリズムを先に説明することとする。また、下にすべての手法でのアルゴリズムを大まかに示す。

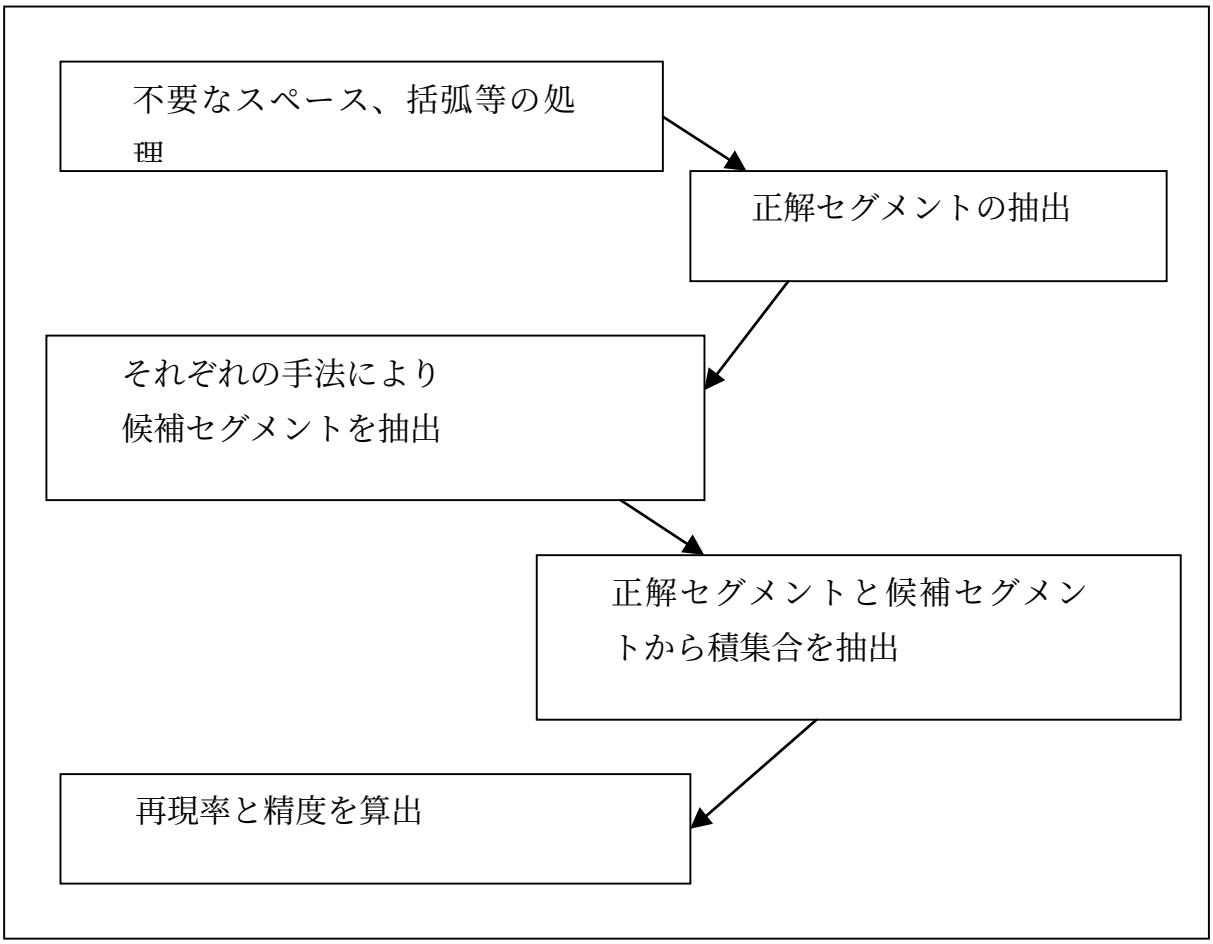

図 3.3　すべての手法で共通するアルゴリズム

まず、不要なスペース、括弧等の処理であるが、これは対話データの文字コード

や半角、全角等が統一していない場合に、主にパターンマッチングを使っているために、マッチすべきところでマッチしない等の不備を防ぐための処理である。

次に正解セグメントの抽出であるが、これは、パターンマッチングで抽出を行なう。

3 番目の、それぞれの手法独自で候補セグメントを抽出する部分は次節より説明することとする。

4 番目の正解セグメントと候補セグメントから積集合を抽出する部分であるが、これは前節で説明した、再現率、精度の算出に必要なために行なう処理である。

最後の再現率と精度の算出であるが、ここでは、前節で述べた式を用いて最終的な評価のための数値を算出するためである。

第 4 章
手がかり語を用いた手法についての評価、考察

本章では、手がかり語を用いた手法を評価し、その評価に基づく考察を行なう

4.1 手がかり語を用いた手法

手がかり語の中でも、談話標識、接続標識、フィラーといったようにそれぞれの語の機能を重視した手がかり語と、接続詞、感動詞、フィラーといったように形態素解析の結果の手がかり語の 2 つがある。本節ではそれぞれの手法について別々に述べることとする。

4.1.1 機能を重視した手がかり語を用いた手法

国語や英語の読解問題等で、「しかし」という接続詞が出てきたら前に書かれていた内容と逆の内容が次からは書かれるなどというように、接続詞をはじめとした手がかり語はある程度の文のまとまりの前後関係に重要な役割を果たしていると考え

られる。そこで、手がかり語を用いた手法として、手がかり語のうちの接続標識、談話標識、フィラーを用いてセグメンテーションを行なう手法について提案する。

まず、この手法で扱う手がかり語のうちの接続標識、談話標識、フィラーのそれぞれの定義であるが、これは、[人工知能学会談話対話におけるコーパス利用研究グループ 99]に従ってタグ付けされたものであり、従来の品詞の定義とは若干異なる部分があり、機能を重視した定義であることを付け加えておく。

接続標識は、[人工知能学会談話対話におけるコーパス利用研究グループ 99]が接続標識のリストとしてあげたものが接続詞として出現した場合につけられたタグで、さらに、接続標識に助詞「ね」、「さ」、「な」、「よ」、「ですね」が後続した場合は後続する助詞を含めて接続標識とするという条件が付け加えられている。

談話標識は、話題の始まり、転換、途切れが会話の再開など、談話同士の対応付けの機能を持つものであるという定義がなされている。

フィラーは、思案・逡巡・発話継続合図などを示す場つなぎの発話であるという定義がなされている。

この手法では、接続標識、談話標識、フィラーのそれぞれにマッチした発話をセグメントの開始部とみなすというものである。以下に例を示す。

[1:教室会議の日時:]

0012 R:{C では}{F えと}月曜日の２時から{F え}３時半まで教室{会議[?]}を登録してください

0013 L:はい

[1:教室会議の日時:]

0014 L:月曜日の１４時から{F え}１５時半まで教室会議を登録します

図 4.1 接続標識、談話標識、フィラーを用いた手法の例
（C：接続標識、D：談話標識、F：フィラーに対応）

この場合、接続標識、談話標識、フィラーのそれぞれにマッチした発話をセグメントの開始部とみなすので、接続詞を手がかりにした場合、0012 がセグメントの候

補となり、フィラーを手がかりにした場合、0012、0014 がセグメントの候補となる。

4.1.2 形態素解析の結果の手がかり語を用いた手法

機能を重視した手がかり語の中でも談話標識を用いたセグメンテーションは有効であるという結果が示されている[中里 99]が、接続標識、談話標識等の機能を重視した手がかり語のタグ付けは現状では自動的には行なえず、人手でタグ付けを行なうしかないため、労力と時間がかかるという問題点がある。

そこで、表層的手がかりを用いてセグメンテーションを行なうため、現状で何らかの表層的な手がかりに関するタグ付けを自動的に行なうことを考えた場合、形態素解析プログラムを使うことで対話データに自動的に品詞情報を付与できることから品詞情報を使うことが有効であると考えられる。

ここで大切なことは、機能を重視した手がかり語の利用はタグ付けに問題がありそのままつかえないため、形態素解析した結果の手がかり語で代用するという考え方である。そこで、前節で利用した機能を重視した手がかり語のうちの接続標識、談話標識、フィラーとの対応関係をとる必要性が生まれてくる。

機能を重視した手がかり語のうちの接続標識は接続詞として、フィラーはそのままフィラーとして品詞情報に直した際にも現れてくる。しかし、談話標識を形態素解析の結果の手がかり語で代用するには、特定の形態素解析した結果の手がかり語を割り当てることは困難であると考えられる。例えば、談話標識の中でも「じゃ」、「では」などのように形態素解析の結果の手がかり語に直すと接続詞に分類されるものもあれば、「といいますと」、「ということは」などのように複数の文節に分かれてしまうものもある。そこで、今回、談話標識は接続詞に含むことで、談話標識の代用とすることにする。さらに、機能を重視した手がかり語には現れないが、形態素解析した結果の手がかり語に現れてくるものとして感動詞がある。感動詞には「じゃ」などといった談話標識に相当する語が含まれている。また、「うん」などといったフィラーに相当する語が含まれている。そこで、この感動詞を談話標識、フィラ

ーの代用として利用することにする。

この手法では、接続詞、感動詞、フィラーの 3 つを利用してセグメンテーションを行なうことにする。具体的には、接続詞、感動詞、フィラーのそれぞれにマッチした箇所を候補セグメントとする。以下に例を示す。

---

[1] o
12 33 R では 接続詞
12 34 R えと フィラー
12 35 R 月曜日 名詞-副詞可能
12 36 R の 助詞-連体化
12 37 R ２ 名詞-数
12 38 R 時 名詞-接尾-助数詞
12 39 R から 助詞-格助詞-一般
12 40 R え フィラー
12 41 R ３ 名詞-数
12 42 R 時半 名詞-接尾-助数詞
12 43 R まで 助詞-副助詞
12 44 R 教室 名詞-一般
12 45 R 会議 名詞-サ変接続
12 46 R を 助詞-格助詞-一般
12 47 R 登録 名詞-サ変接続
12 48 R する 動詞-自立
12 49 R て 助詞-接続助詞
12 50 R くださる 動詞-非自立
EOS
[] -
13 52 L はい 感動詞
EOS
[1] x
14 53 L 月曜日 名詞-副詞可能

14 54 L の 助詞-連体化
14 55 L １４ 名詞-数
14 56 L 時 名詞-接尾-助数詞
14 57 L から 助詞-格助詞-一般
14 58 L え フィラー
14 59 L １５ 名詞-数
14 60 L 時半 名詞-接尾-助数詞
14 61 L まで 助詞-副助詞
14 62 L 教室 名詞-一般
14 63 L 会議 名詞-サ変接続
14 64 L を 助詞-格助詞-一般
14 65 L 登録 名詞-サ変接続
14 66 L する 動詞-自立
14 67 L ます 助動詞
EOS

図 4.1.2 接続詞、感動詞、フィラーを用いた手法の例

この場合、接続詞、感動詞、フィラーのそれぞれにマッチした発話をセグメントの開始部とみなすので、接続詞を手がかりにした場合、12 がセグメントの候補となり、感動詞を手がかりにした場合、13 がセグメントの候補となり、フィラーを手がかりにした場合、0012、0014 がセグメントの候補となる。

ここで、前節と同様にフィラーの本来もつ意味合いを考え、フィラーをセグメントの終了箇所とみなす。また、上に述べたとおり感動詞のうち、フィラーに相当する語が含まれている可能性があるため、感動詞もセグメントの終了箇所とみなし、接続詞のみを 1 つ目の方法と同様にセグメントの開始箇所とみなすことにする。これが 2 つ目の方法である。

上の図 3 で説明すると、接続詞を手がかりにした場合、1 つ目の方法と同様に 12 がセグメントの候補となる。感動詞を手がかりにした場合、13 をセグメントの終了箇所とみなすため、14 がセグメントの候補となる。フィラーを手がかりにした場合、

12、14 をセグメントの終了箇所とみなすため、それぞれの発話の次の発話である、13、15 がセグメントの候補となる。

4.2 手がかり語を用いたセグメンテーションの方法

手がかり語を用いた評価実験では、前節で上げたとおり、機能を重視した手がかり語の接続標識、談話標識、フィラーの 3 つのタグ、また、形態素解析の結果の手がかり語の接続詞、感動詞、フィラーの 3 つのタグに対するパターンマッチングで、3 つそれぞれにマッチする箇所の発話番号を候補セグメントとした。

4.3 手がかり語を用いた手法の結果と考察

まず、機能を重視した手がかり語の接続標識、談話標識、フィラーに関するタグ付けをしてあるデータが 14 対話分しか存在しないため、14 対話のみの結果のみ算出した。接続標識、談話標識、フィラーそれぞれのみの結果を表 4.3.1 に示す。なお、C は接続標識、D は談話標識、F はフィラーにそれぞれ対応する。

	C		D		F	
	再現率	**精度**	**再現率**	**精度**	**再現率**	**精度**
99atr01	5.6%	33.3%	22.2%	100.0%	22.2%	40.0%
99atr02	13.6%	100.0%	31.8%	87.5%	45.5%	66.6%
99chi0a	12.2%	100.0%	38.8%	95.0%	26.5%	81.3%
99crl01	23.8%	100.0%	4.8%	50.0%	42.9%	81.8%
99crl02	9.7%	100.0%	9.7%	75.0%	32.3%	62.5%
99crl03	8.3%	100.0%	5.6%	100.0%	27.8%	76.9%
99kyo01	0.0%	0.0%	25.0%	100.0%	75.0%	63.2%

99kyo0a	4.0%	100.0%	32.0%	100.0%	52.0%	61.9%
99osa01	3.8%	33.3%	26.9%	87.5%	84.6%	56.4%
99osa02	6.0%	50.0%	19.7%	86.7%	78.8%	67.5%
99osa0a	8.7%	100.0%	26.1%	85.7%	34.8%	50.0%
99tsu0a	0.0%	0.0%	6.5%	40.0%	3.2%	7.7%
99uec0a	4.2%	50.0%	20.8%	83.3%	50.0%	40.0%
99was03	0.0%	0.0%	20.5%	100.0%	15.4%	50.0%
AVE.	7.1%	61.9%	20.7%	85.1%	42.2%	57.6%

表 4.3.1　接続標識、談話標識、フィラーを用いた結果

この結果では、談話標識の精度の平均が 85.1%と非常に高い結果を示しており、談話標識があればセグメントであると考えることが有効であると言える。また、接続標識の精度も平均が 61.9%とそれほどでも高い数値ではないが、接続標識の再現率と合わせて考えると接続標識自体の数は少ないが、接続標識があるところでは精度が高くなっていることから、接続標識があればセグメントと考えることが有効であると言える。

以上の結果を推測すると、接続標識、談話標識をセグメンテーションに利用することが有効であると考えられるが、フィラーをセグメントの開始として利用することが有効ではないということが言える。

次に、接続詞、感動詞、フィラーそれぞれのみの 14 対話分の結果を表 4.3.2 に示す。なお、Ｃは接続詞、Ｄは感動詞、Ｆはフィラーにそれぞれ対応する。

	C		D		F	
	再現率	**精度**	**再現率**	**精度**	**再現率**	**精度**
99atr01	29.4%	71.4%	5.9%	5.6%	17.6%	30.0%
99atr02	40.0%	72.7%	20.0%	26.7%	45.0%	60.0%
99chi0a.	60.0%	65.2%	4.0%	1.4%	32.0%	50.0%
99crl01.	36.4%	57.1%	9.1%	3.0%	54.5%	54.5%
99crl02.	15.8%	42.9%	21.1%	16.0%	36.8%	43.8%
99crl03.	15.4%	100.0%	15.4%	10.5%	34.6%	69.2%
99kyo01.	23.1%	100.0%	0.0%	0.0%	84.6%	57.9%

99kyo0a.	33.3%	66.7%	0.0%	0.0%	55.6%	47.6%
99osa01.	33.3%	72.7%	4.2%	2.0%	87.5%	53.8%
99osa02.	22.4%	72.2%	12.1%	5.4%	79.3%	59.7%
99osa0a.	35.3%	66.7%	5.9%	20.0%	23.5%	25.0%
99tsu0a.	6.7%	40.0%	10.0%	6.4%	3.3%	7.7%
99uec0a.	20.0%	80.0%	10.0%	5.1%	55.0%	36.7%
99was03.	21.1%	100.0%	15.8%	15.0%	15.8%	50.0%
AVE.	28.0%	72.0%	9.5%	8.4%	44.7%	46.1%

表 4.3.2 接続詞、感動詞、フィラーを用いたセグメンテーションの結果 1

この結果によると、接続詞の精度が 72%と若干高い数値をしめしている他は高い数値を示しているものは見られなかった。また逆に感動詞の精度が 8.4%と非常に低い数値を示している。接続詞に関して言えば精度の割には再現率が低いことから接続詞があればセグメントになる可能性が若干高いという推測ができる。ここで、36 対話について同様の結果を表 4.3.3 に示す

	C		D		F	
	再現率	**精度**	**再現率**	**精度**	**再現率**	**精度**
2-04-0.	47.7%	82.4%	10.2%	20.9%	22.7%	37.7%
2-04-1.	33.8%	69.4%	17.6%	23.2%	31.1%	69.7%
2-04-2.	39.2%	80.9%	19.6%	29.7%	19.6%	50.0%
2-05-0.	33.3%	73.7%	16.7%	22.6%	23.8%	52.6%
2-05-1.	16.0%	53.3%	42.0%	52.5%	32.0%	72.7%
2-05-2.	34.2%	65.0%	21.1%	32.0%	21.1%	47.1%
2-08-0.	24.6%	68.2%	23.0%	41.2%	52.5%	68.1%
2-08-1.	24.5%	61.9%	18.9%	31.3%	45.3%	54.5%
2-08-2.	18.6%	66.7%	14.0%	23.1%	44.2%	65.5%
2-10-0.	30.2%	64.0%	22.6%	15.8%	66.0%	55.6%
2-10-1.	32.1%	70.8%	30.2%	21.6%	67.9%	65.5%
2-10-2.	31.6%	81.8%	29.8%	20.5%	70.2%	64.5%
2-12-0.	38.9%	67.7%	22.2%	26.7%	27.8%	48.4%
2-12-1.	37.9%	86.2%	19.7%	26.0%	39.4%	66.7%
2-12-2.	42.3%	66.7%	23.1%	24.0%	25.0%	61.9%
2-14-0.	35.9%	82.4%	12.8%	15.6%	46.2%	66.7%

2-14-1.	33.3%	69.0%	6.9%	9.5%	47.1%	70.7%
2-14-2.	32.7%	85.0%	13.5%	13.7%	57.7%	78.9%
2-15-0.	35.4%	73.9%	35.4%	30.9%	43.8%	53.8%
2-15-1.	27.9%	70.4%	22.1%	26.3%	33.8%	60.5%
2-15-2.	34.5%	83.3%	24.1%	26.9%	43.1%	54.3%
2-17-0.	48.3%	82.4%	17.2%	31.3%	37.9%	61.1%
2-17-1.	48.5%	69.6%	19.7%	28.3%	19.7%	61.9%
2-17-2.	40.0%	76.2%	25.0%	28.6%	30.0%	42.9%
2-18-0.	29.0%	73.0%	20.4%	18.6%	31.2%	64.4%
2-18-1.	37.5%	70.6%	18.8%	17.4%	37.5%	66.7%
2-18-2.	23.0%	69.0%	20.7%	18.4%	37.9%	63.5%
2-19-0.	37.5%	70.6%	26.6%	27.0%	40.6%	61.9%
2-19-1.	29.3%	63.0%	20.2%	20.0%	34.3%	56.7%
2-19-2.	36.8%	71.4%	20.6%	24.6%	33.8%	71.9%
2-20-0.	17.4%	70.6%	21.7%	21.7%	58.0%	67.8%
2-20-1.	32.3%	74.1%	29.0%	26.5%	62.9%	68.4%
2-20-2.	30.8%	76.2%	13.5%	12.7%	51.9%	69.2%
2-23-0.	37.5%	58.3%	17.9%	17.2%	50.0%	65.1%
2-23-1.	20.8%	65.2%	22.2%	24.6%	52.8%	67.9%
2-23-2.	22.2%	70.0%	20.6%	21.0%	50.8%	69.6%
AVE.	32.7%	71.7%	21.1%	24.2%	41.4%	61.8%

表 4.3.3　接続詞、感動詞、フィラーを用いたセグメンテーションの結果 2

36 対話についての接続詞、感動詞、フィラーそれぞれの結果は、接続詞の精度が71.7%と高い数値が得られた。また、14 対話のほうでは非常に低い数値を示していた感動詞が、61.8%まで高くなっており、14 対話と比べると全体的に高い数値を示している。

ここで問題になるのは、まず、機能面を重視した手がかり語である接続標識、談話標識、フィラーと形態素解析の結果の手がかり語である接続詞、感動詞、フィラーとの対応である。機能面を重視した手がかり語は接続標識、談話標識、フィラーを用い、形態素解析の結果の手がかり語では接続詞、感動詞、フィラーを利用してきた。機能面を重視した手がかり語の結果は 14 対話分のみであるので、機能面を重視したの 14 対話分と形態素解析の結果の手がかり語である接続詞、感動詞、フィラーの 14 対話分を比較すると、表面上同じと思われる機能面を重視した手がかり語の

接続標識と形態素解析の結果の手がかり語の接続詞の結果の精度は 10%程度、形態素解析の結果の手がかり語のほうが高い。そこで、14 対話中の機能面を重視した手がかり語である接続標識と形態素解析の結果の手がかり語である接続詞の分析を行なった。

その結果、機能面を重視した手がかり語のタグ付けでは談話標識とされていたものが、形態素解析の結果の手がかり語では接続詞にタグ付けされるものがあることがわかった。このことから、機能面を重視した手がかり語では談話標識とされていたものが、形態素解析の結果の手がかり語の場合に接続詞とされるため、精度が向上したと推測できる。

次に問題となるのが、形態素解析の結果の手がかり語のフィラーについて、14 対話と 36 対話を比較すると 36 対話のフィラーの精度が 20%弱向上している点である。そこで、形態素解析の結果の手がかり語のフィラーに関して 14 対話と 36 対話との比較分析を行なった。

その結果、14 対話データでのフィラーと 36 対話データでのフィラーの内訳には大きな違いが見られなかったが、14 対話のほうでフィラーとされている「そうですね」といった複数の文節から構成される語が 36 対話データではフィラーとはされず、文節に分けられているという特徴が見られた。他に 36 対話では「うん」がフィラーに分類され、この「うん」は応答の働きをもつものであり、セグメントの候補になりうるものであるので、こういった応答的な役割を果たす語が 36 対話でフィラーに含まれているために、精度が向上したことが要因として上げられる。

第 5 章
実質的に意味のない語を用いた手法と評価、考察

本章では、[金 00]の実質的に意味のない語を用いた手法を改良したものを実験、評価し、その評価に基づく考察を行なう。

5.1 実質的に意味のない語を用いた手法

我々が人と話をしているときに、話の内容が確実に変わったことを認識するまでには、何ターンかのやりとりが必要である。その何ターンかのやりとりは、「うんうん」や「はいはい」などと同意を表わすものであったり、「ありがとうございました」や「お願いします」などの挨拶表現であったり、具体的に話されている内容とは直接的に関係があるとは直感的には認識できないような語で成り立っていると考えられる。

それに関連した手法として、[金 00]では、対話データから直感的に実質的に意味をもたないと思われる発話を抽出し、それを実質的に意味のない語として、それをもとに人手でセグメンテーションを行なっている。今回は、[金 00]で抽出された実質的に意味のない語を用いてセグメンテーションを行なう手法を提案する。なお、ここで利用する実質的に意味のない語のリストの一部は前出の図 2.3 に示した。

この手法では、[金 00]の実質的に意味のない語以外を候補セグメントとする手法が直感に合わないと判断されたため、実質的に意味のない語をセグメントの終了部とみなす手法を提案する。以下より図 5.1 の例をもとに説明する。

0010 R:{ちょ[?]}
0011 R:(分)かりました
[1:教室会議の日時:]
0012 R:{C では}{F えと}月曜日の２時から{F え}３時半まで教室{会議[?]}を登録してください
0013 L:はい
[1:教室会議の日時:]
0014 L:月曜日の１４時から{F え}１５時半まで教室会議を登録します

図 5.1　実質的に意味のない語を用いた手法の例

この場合、実質的に意味のない語のリストにマッチする発話が 0010、0011、0013 にあり、実質的に意味のない語にマッチした発話の次をセグメントの開始部と見なすので、それぞれの次の発話である、0011、0012、0014 がセグメントの候補となる。

さらに、この手法において、連続してセグメントの候補となった箇所がある場合は、連続してマッチした箇所のうち、マッチした箇所をセグメントの終了箇所と見なすので、最も後にマッチした箇所をセグメントの候補を考える際の対象にするというアルゴリズムを提案する。

例えば、図 5.1 を用いて説明すると、実質的に意味のない語にマッチする発話は、0010、0011、0013 であり、実質的に意味のない語にマッチした箇所がセグメントの終了部と見なすので、連続してマッチしている 0010、0011 の部分に関しては、0011 がマッチしたと見なされ、0012 がセグメントの候補となる。また、0013 に関しては、連続してマッチしているわけではないので、そのまま 0014 がセグメントの候補となる。

5.2 実質的に意味のない語を用いたセグメンテーションの方法

実質的に意味のない語を用いた評価実験では、前節で上げた方法で、2.3 節の図 2.3 に上げた実質的に意味にない語リストと発話データとのパターンマッチングで、リスト内の語が発話データにマッチする発話番号を抜き出し、実質的に意味のない語をセグメントの終了部と見なし、マッチした発話番号の次の発話番号を候補セグメントとした。

5.3 実質的に意味のない語を用いた手法の結果と考察

まず、実質的に意味のない語を用いた手法の 14 対話分の結果を表 5.3.1 に示す。なお、ここで用いている実質的に意味のない語自体は[金 00]において、この 14 対話からリストアップしたものを用いている。

表 5.3.1 は、[金 00]では、実質的に意味のない語以外にマッチした発話をセグメントの候補としており、実質的に意味があるような発話すべてがセグメントの候補になっていた。

そこで本研究では、実質的に意味のない語があった場合、実質的に意味のない語の次の発話から新しい話題になる可能性があると考えられ、実質的に意味のない語がマッチした発話を今まで話されてきた話題の終わりとみなし、実質的に意味のない語にマッチした発話の次の発話をセグメントの候補とした。

	kim2_1.pl	
	再現率	**精度**
99atr01.	50.0%	69.2%
99atr02.	45.5%	66.7%
99chi0a.	73.5%	40.9%
99crl01.	71.4%	41.7%
99crl02.	51.6%	61.5%
99crl03.	55.6%	51.3%
99kyo01.	50.0%	66.7%
99kyo0a.	56.0%	63.6%
99osa01.	84.6%	45.8%
99osa02.	74.2%	40.5%
99osa0a.	26.1%	75.0%
99tsu0a.	90.3%	62.2%
99uec0a.	83.3%	48.8%
99was03.	46.2%	60.0%
AVE.	61.3%	56.7%

表 5.3.1　実質的に意味のない語を用いたセグメンテーションの結果 1

表 5.3.1 の結果では、再現率が 61.3%、精度が 56.7%という結果を得た。再現率が若干高いことから、セグメントの候補の数自体が多い可能性があるため、セグメントの候補数を減らすことを考えた場合、この方法では、実質的に意味のない語にマッチした発話の次の発話すべてをセグメントの候補としているので、もし、実質的に意味のない語が複数回連続して続いた場合でも、複数回分マッチしたとみなしていることになり、セグメントの候補も増えることが考えられる。そこで、実質的に意味のない語が複数回連続して続いた場合は、複数回連続した発話のうちの一番最後の発話の次の発話をセグメントの候補とした。その結果を表 5.3.2 に示す。

表 5.3.2　実質的に意味のない語を用いたセグメントの結果 2

	kim2_2.pl	
	再現率	**精度**
99atr01.	50.0%	69.2%
99atr02.	45.5%	66.7%
99chi0a.	73.5%	90.0%
99crl01.	66.7%	87.5%
99crl02.	51.6%	80.0%
99crl03.	55.6%	80.0%
99kyo01.	50.0%	100.0%
99kyo0a.	56.0%	87.5%
99osa01.	84.6%	81.5%
99osa02.	72.7%	78.7%
99osa0a.	26.1%	85.7%
99tsu0a.	90.3%	80.0%
99uec0a.	79.2%	86.4%
99was03.	46.2%	85.7%
AVE.	60.6%	82.8%

この結果では、前の結果よりも精度が 20%強よくなる結果が示された。これは、実質的に意味のない語が複数以上連続した場合に、その連続した部分は、実質的に意味のないまとまりになり、実質的に意味のない語自体が、実質的に意味があると考えられる発話には関係がないということになり、直感に合う結果であると言える。

次に、同様の 2 つの方法を 36 対話に適用した、この場合も 14 対話の時と同様に実質的に意味のない語は 14 対話からリストアップした語を用いた。表 5.3.3 に結果を示す。なお、表 5.3.3 の kimseg2_1 は実質的に意味のない語にマッチした発話の次の発話番号をセグメントの候補とした結果を、kimseg2_2 には実質的に意味のない語が複数回連続してマッチした場合には、複数回マッチした発話のうちの一番最後の発話の次の発話をセグメントの候補とした結果を示す。

kimseg2_1	kimseg2_2

	再現率	精度	再現率	精度
2-04-0.	39.8%	62.5%	39.8%	68.6%
2-04-1.	35.1%	59.1%	35.1%	60.5%
2-04-2.	30.9%	62.5%	30.9%	69.8%
2-05-0.	28.6%	52.2%	28.6%	57.1%
2-05-1.	52.8%	65.1%	52.8%	68.3%
2-05-2.	47.4%	48.6%	44.7%	51.5%
2-08-0.	26.2%	59.3%	26.2%	64.0%
2-08-1.	31.5%	36.2%	31.5%	37.8%
2-08-2.	39.5%	50.0%	39.5%	50.0%
2-10-0.	61.1%	52.4%	59.3%	56.1%
2-10-1.	52.8%	36.8%	52.8%	41.8%
2-10-2.	62.1%	43.9%	60.3%	46.7%
2-12-0.	42.6%	52.3%	38.9%	55.3%
2-12-1.	42.4%	46.7%	42.4%	49.1%
2-12-2.	51.9%	52.9%	51.9%	57.4%
2-14-0.	59.0%	56.1%	59.0%	57.5%
2-14-1.	44.8%	46.4%	44.8%	48.8%
2-14-2.	56.6%	44.1%	56.6%	50.0%
2-15-0.	47.9%	47.9%	43.8%	55.3%
2-15-1.	44.1%	62.5%	39.7%	64.3%
2-15-2.	48.3%	53.8%	48.3%	65.1%
2-17-0.	27.6%	72.7%	27.6%	80.0%
2-17-1.	30.3%	52.6%	28.8%	51.4%
2-17-2.	52.5%	72.4%	50.0%	74.1%
2-18-0.	52.7%	50.5%	51.6%	57.1%
2-18-1.	47.6%	41.7%	47.6%	45.5%
2-18-2.	50.6%	43.1%	48.3%	46.2%
2-19-0.	48.5%	55.2%	42.4%	53.8%
2-19-1.	45.5%	54.9%	42.4%	59.2%
2-19-2.	36.8%	59.5%	35.3%	63.2%
2-20-0.	58.6%	56.2%	55.7%	58.2%
2-20-1.	56.5%	42.7%	54.8%	50.0%
2-20-2.	61.5%	59.3%	61.5%	65.3%
2-23-0.	48.2%	49.1%	46.4%	57.8%
2-23-1.	38.4%	51.9%	35.6%	55.3%
2-23-2.	60.3%	56.7%	60.3%	65.5%
AVE.	46.1%	53.0%	44.9%	57.1%

表 5.3.3　実質的に意味のない語を用いたセグメントの結果 3

表 5.3.3 の結果では、どちらの場合も再現率が 50%弱で、精度が 50%強という結果を得た。表 5.3.1 および表 5.3.2 の 14 対話を用いたものと比較すると結果が悪くなっている。これは、14 対話を用いたものでは、実質的に意味のない語のリストを 14 対話そのものから作成し使用しているが、36 対話のほうでは、36 対話そのものから作成したリストではなく、14 対話から作成したリストを使っていることが要因であると考えられる。また、ここから、14 対話から作成した実質的に意味のない語のリストが汎用的ではなく、個別のタスクにしか対応しない可能性があることが考えられる。

そこで、14 対話から作成した実質的に意味のない語が 36 対話でどういうものがどれくらい対応しているかを分析した。14 対話から作成したリストに実質的に意味のない語として上げられているものは全部で 112 語あり、36 対話でそのリストとマッチした語は 25 語であった。その内訳を図 5.3 に示す。

「はい」、「うん」、「そうですか」、「そう」、「ふんふん」、「あ」、「はいはい」、「そうです」、「なるほど」、「ああはい」、「笑い」、「ええ」、「ありがとうございました」、「あそうですね」、「あそうですか」、「そうそうそう」、「はいはいはい」、「はいそうですか」、「ああそうですか」、「あはい」、「うんうんうん」、「うんうん」、「うんはい」、「ね」、「はあはあはあはあ」

図 5.3　36 対話でマッチした実質的に意味のない語リスト

図 5.3 の特徴として、応答や同意を表す語が多い。ここから、実質的に意味のない語が応答、同意などの機能を持つ可能性が高いと考えられる。

第 6 章
テキストタイリングアルゴリズムの手法と評価、考察

本章では、テキストタイリングアルゴリズムを用いた手法を評価し、その評価に基づく考察を行なう。

6.1 テキストタイリングアルゴリズムを用いた手法

直感的に考えると、我々が日常において会話をしている際に使われる単語、特に名詞や動詞に関わる語の違いで意味を読み取ることが多いと考えられる。また、話題の変化も会話中に使われるそれらの単語の違いによって察知することが直感的ではあるが推測できる。前節までは手がかり語や接続詞、感動詞、フィラー、実質的に意味のない語といった表層的手がかりを用いてきたが、本節では、会話中の単語の分布に着目した内容的手がかりを用いていくこととする。そこで、文書中の単語の分布に着目したテキストタイリングアルゴリズムを対話に適用し、セグメンテーションを行なう手法について提案する。

テキストタイリングアルゴリズムを用いた手法では、まず、1 ブロック内の発話単位を決める必要がある。なお、[Hearst 99]では、1 ブロック内の単位を 2 トークン列と設定し、1 トークンを 20 語と設定している。

本論文では、ブロック間の索引語頻度を用いてセグメンテーションを行なう手法と、隣り合うブロック内の新出単語の頻度を用いてセグメンテーションを行なう手法の 2 つを実験、評価する。

まず、1 つ目のブロック間の索引語頻度を用いてセグメンテーションを行なう手法であるが、手順としては、まず、ブロック内の発話数を設定する必要がある。この場合、隣接ブロック間の索引語頻度をもとにした類似度を算出する必要があるため、設定した発話数でブロックを区切り、さらに、隣接するブロックを設定してあげる必要がある。下にブロックの設定、および隣接するブロックの設定について例を上げる。

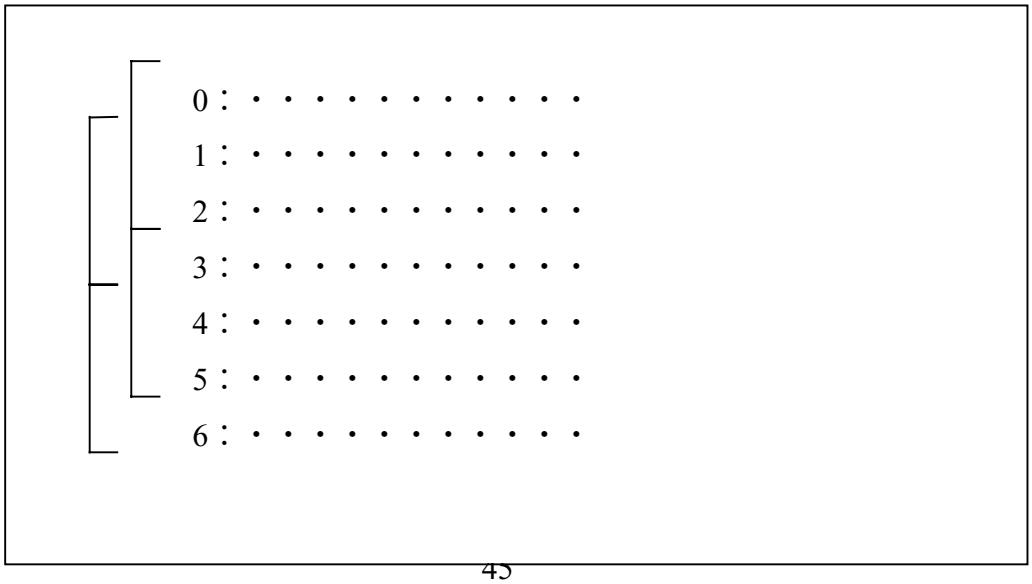

図 6.1.1　ブロックの設定、および隣接するブロックの設定について例

図 6.1 について、左側の数字は発話番号を表わすものとする。まず、1 ブロック内の発話数を 3 とした場合、1 番目のブロックは 0〜2 番目までの発話、2 番目のブロックは 3〜5 番目までの発話で構成される。最終的には隣接するブロック同士で計算をするため、1 番目のブロックと 2 番目のブロックは 1 セットとして考える。次のブロックの設定をする場合は、発話番号が 1 番目のものからはじまり、1 番目のブロックは 1〜3 番目までの発話、2 番目のブロックは 4〜6 番目までの発話で構成され、隣接ブロックは 1〜3 番目までの発話で構成されるブロックと 4〜6 番目までの発話で構成されるブロックが 1 セットということになる。ここで、最後のブロックおよび、隣接ブロックのセットが設定した発話数通りにとってくることができない場合について触れておくと、この場合は、一番最後の隣接ブロックのセットが設定した 1 ブロック中の発話数分を確保できない箇所はブロック設定を行なわないようにする。

次に、ブロックおよび、隣接ブロックのセットの設定が行なわれた後の処理についてであるが、隣接ブロックのセット中のそれぞれのブロック内での単語の頻度を算出し、それぞれのブロックの類似度を計算する。この際の計算式は、図 6.1.2 のようになる。

ただし、w はそれぞれのブロックでの単語の頻度を表わし、b1、b2 はそれぞれのブロックを表わし、t はそれぞれのブロック内の個々の単語を表わす。

$$score(i) = \frac{\Sigma_t w_{t,b1} w_{t,b2}}{\sqrt{\Sigma_t w^2_{t,b1} \Sigma_t w^2_{t,b2}}}$$

図 6.1.2　類似度の計算式

さらに、計算した類似度を平滑化し、それをもとに、対話データ全体での平均と標準偏差を算出し、図 6.1.3 の条件を満たすものをセグメントの候補とする。ただし、$\bar{s}$ は平均を表わし、σ は標準偏差を表わし、c は定数を表わし、c =0.5 かc =1.0 をと

るように設定する。*smoothingscore*(i) は i 番目における平滑化された類似度を表わす。

$$\bar{s} - c\,\sigma \geq smoothingscore(i)$$

図 6.1.3　セグメント決定の条件式

次に、2 つ目の隣り合うブロック内の新出単語の頻度を用いてセグメンテーションを行なう手法であるが、手順としては、1 つ目のブロック間の索引語頻度を用いてセグメンテーションを行なう手法と同様に、まず、ブロック内の発話数を設定する必要がある。次に隣接ブロック内での前のブロックに対する後のブロックの新出単語の頻度と後のブロックに対する前のブロックの新出単語の頻度と、それぞれの新出単語の総計を算出し、それをもとに類似度を算出する。この際の計算式は図 6.1.4 のようになる。ただし、b1、b2 はそれぞれのブロック、*NumNewTerms* はそれぞれのブロックに対する新出単語の頻度を表わし、*TotalNewTerms* は隣接ブロック内での新出単語の総計を表わす。

$$score(i) = \frac{NumNewTerms(b1) + NumNewTerms(b2)}{TotalNewTerms(b1, b2)}$$

図 6.1.4　新出単語の頻度を用いてセグメンテーションを行なう手法の類似度

さらに、1 つ目のブロック間の索引語頻度を用いてセグメンテーションを行なう手法と同様に、計算した類似度を平滑化し、それをもとに、対話データ全体での平均と標準偏差を算出し、図 3.4.4 の条件を満たすものをセグメントの候補とする。

6.2 テキストタイリングアルゴリズムを用いた

方法

テキストタイリングアルゴリズムを用いた実験、評価では、前節で上げたとおり、まず、ブロック数を設定した発話数ごとに区切る作業を行なう。その上で、隣接するブロックの索引語頻度を算出する方法と、隣接するブロックの新出単語をカウントする方法の 2 つを行なう。

6.3 テキストタイリングアルゴリズムを用いた手法の結果と考察

テキストタイリングアルゴリズムを用いた手法では、まず、隣接ブロック間の索引語頻度をもとに境界算出の定数を 1.0 として算出した類似度からセグメンテーションを行なう手法の 14 対話の結果を図 6.3.1 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	精度

99atr01.	64.7%	34.4%	76.5%	40.6%	76.5%	41.9%
99atr02.	75.0%	45.5%	75.0%	48.4%	65.0%	46.4%
99chi0a.	88.0%	18.0%	72.0%	15.0%	72.0%	15.7%
99crl01.	91.7%	22.4%	91.7%	22.4%	83.3%	21.7%
99crl02.	85.0%	30.9%	85.0%	33.3%	75.0%	31.9%
99crl03.	88.9%	35.3%	74.1%	32.8%	66.7%	31.0%
99kyo01.	69.2%	33.3%	76.9%	35.7%	53.8%	28.0%
99kyo0a.	83.3%	33.3%	77.8%	34.1%	66.7%	29.3%
99osa01.	91.7%	26.8%	79.2%	25.7%	79.2%	26.4%
99osa02.	91.4%	28.2%	94.8%	29.6%	89.7%	28.9%
99osa0a.	76.5%	36.1%	70.6%	34.3%	64.7%	30.6%
99tsu0a.	83.3%	26.6%	83.3%	29.8%	76.7%	26.7%
99uec0a.	70.0%	22.6%	85.0%	28.3%	75.0%	26.3%
99was03.	81.6%	51.7%	76.3%	50.0%	76.3%	50.0%
AVE.	81.4%	31.8%	79.9%	32.9%	72.9%	31.1%

表 6.3.1　テキストタイリングアルゴリズムでのセグメンテーションの結果 1

この結果によると、１ブロック数の設定発話数がいずれの場合においても、同じような数値が算出されており、よい結果を示していないことは明らかである。また、再現率に注目すると、どの発話数でも 8 割前後とある程度高く、候補となるセグメントの数が非常に多くなっていることが分かる。次に同様の 36 対話の定数 1.0 の時の結果を表 6.3.2 に示す。

	ブロック２		ブロック３		ブロック４	
	再現率	精度	再現率	精度	再現率	精度
2-04-0.	79.5%	47.0%	77.3%	46.3%	75.0%	46.5%
2-04-1.	78.4%	46.8%	83.8%	50.8%	81.1%	50.4%
2-04-2.	83.5%	49.4%	83.5%	49.1%	78.3%	47.2%
2-05-0.	73.8%	45.6%	76.2%	46.4%	73.8%	46.3%
2-05-1.	72.0%	48.0%	72.0%	48.6%	70.0%	47.9%
2-05-2.	76.3%	44.6%	81.6%	44.9%	71.1%	43.5%
2-08-0.	82.0%	50.5%	83.6%	52.6%	80.3%	51.6%
2-08-1.	79.2%	44.2%	79.2%	47.2%	81.1%	46.2%
2-08-2.	74.4%	45.7%	72.1%	43.1%	69.8%	44.1%

2-10-0.	84.9%	40.2%	77.4%	37.3%	81.1%	39.1%
2-10-1.	79.2%	42.9%	84.9%	44.6%	81.1%	44.8%
2-10-2.	84.2%	42.9%	87.7%	43.1%	80.7%	41.4%
2-12-0.	77.8%	46.2%	77.8%	47.2%	79.6%	48.9%
2-12-1.	80.3%	52.0%	78.8%	50.5%	81.8%	52.4%
2-12-2.	82.7%	49.4%	88.5%	51.7%	84.6%	51.8%
2-14-0.	76.9%	43.5%	76.9%	45.5%	74.4%	45.3%
2-14-1.	79.3%	46.6%	82.8%	49.3%	80.5%	50.4%
2-14-2.	80.8%	43.3%	88.5%	46.5%	82.7%	46.7%
2-15-0.	83.3%	40.8%	75.0%	37.1%	75.0%	36.7%
2-15-1.	75.0%	45.5%	80.9%	48.7%	80.9%	48.2%
2-15-2.	77.6%	42.1%	81.0%	43.1%	84.5%	45.8%
2-17-0.	69.0%	50.0%	75.9%	56.4%	75.9%	55.0%
2-17-1.	80.3%	49.1%	77.3%	50.0%	77.3%	50.0%
2-17-2.	77.5%	47.0%	80.0%	46.4%	77.5%	46.3%
2-18-0.	86.0%	45.5%	81.7%	43.7%	86.0%	44.7%
2-18-1.	87.5%	45.2%	81.3%	42.6%	84.4%	42.5%
2-18-2.	79.3%	41.6%	86.2%	43.4%	79.3%	42.3%
2-19-0.	78.1%	40.3%	82.8%	42.7%	76.6%	42.2%
2-19-1.	87.9%	39.7%	78.8%	37.0%	85.9%	39.5%
2-19-2.	82.4%	48.3%	82.4%	47.9%	82.4%	47.9%
2-20-0.	81.2%	47.9%	84.1%	47.2%	79.7%	46.6%
	79.5	46.7				
2-20-1. %	%		79.0%	48.0%	80.6%	47.6%
2-20-2.	86.5%	47.9%	80.8%	46.2%	80.8%	44.7%
2-23-0.	82.1%	43.8%	80.4%	44.6%	83.9%	46.5%
2-23-1.	80.6%	46.8%	79.2%	47.1%	72.2%	44.4%
2-23-2.	85.7%	48.2%	77.8%	47.6%	77.8%	41.2%
AVE.	80.2%	45.7%	80.5%	46.2%	79.1%	46.0%

表 6.3.2　テキストタイリングアルゴリズムでのセグメンテーションの結果 2

36 対話の結果においても、14 対話の時と同様に精度はあまり高くなく、再現率が 8 割近いという結果を得た。36 対話の場合においてもセグメントの候補の数が多すぎる可能性が考えられる。以上から推測すると、どのブロックの類似度とも似たような値であるために、候補のセグメントの数が非常に多くなっている可能性が考えられる。

また、隣接ブロック間の索引語頻度をもとに境界算出の定数を 0.5 として算出した類似度からセグメンテーションを行なう手法の 14 対話の結果を表 6.3.3 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	精度
99atr01.	58.8%	34.5%	70.6%	40.0%	64.7%	42.3%
99atr02.	65.0%	44.8%	45.0%	37.5%	55.0%	44.0%
99chi0a.	68.0%	18.7%	52.0%	13.4%	52.0%	13.8%
99crl01.	75.0%	21.4%	83.3%	23.3%	66.7%	21.1%
99crl02.	65.0%	33.3%	70.0%	32.6%	70.0%	34.1%
99crl03.	66.7%	34.6%	70.4%	36.5%	59.3%	30.8%
99kyo01.	61.5%	32.0%	61.5%	33.3%	46.2%	27.3%
99kyo0a.	66.7%	32.4%	72.2%	35.1%	55.6%	28.6%
99osa01.	75.0%	30.5%	66.7%	25.4%	66.7%	25.4%
99osa02.	70.7%	28.3%	75.9%	30.3%	75.9%	31.4%
99osa0a.	64.7%	34.4%	58.8%	33.3%	64.7%	35.5%
99tsu0a.	46.7%	21.2%	56.7%	24.3%	63.3%	26.4%
99uec0a.	65.0%	23.2%	70.0%	26.9%	65.0%	25.5%
99was03.	68.4%	50.0%	71.1%	51.9%	73.7%	53.8%
AVE.	65.5%	31.4%	66.0%	31.7%	62.8%	31.4%

表 6.3.3　テキストタイリングの手法を用いた結果 3

この結果によると、１ブロック数の設定発話数がいずれの場合においても、同じような数値が算出されており、よい結果を示してはいないことが明らかである。定数が 0.5 の時よりも再現率は 20%弱下がっている。ここからセグメントの候補の数が若干少なくなったと推測できる。次に同様の 36 対話の定数 1.0 の時の結果を表 6.3.4 に示す。

	ブロック 2	ブロック 3	ブロック 4

	再現率	精度	再現率	精度	再現率	
2-04-0.	62.5%	47.8%	64.8%	46.7%	63.6%	46.3%
2-04-1.	62.2%	46.5%	62.2%	47.4%	63.5%	50.5%
2-04-2.	72.2%	50.7%	67.0%	47.4%	60.8%	45.7%
2-05-0.	59.5%	46.3%	66.7%	48.3%	66.7%	47.5%
2-05-1.	64.0%	50.0%	62.0%	51.7%	56.0%	47.5%
2-05-2.	57.9%	46.8%	65.8%	43.1%	60.5%	41.1%
2-08-0.	73.8%	54.2%	63.9%	49.4%	63.9%	49.4%
2-08-1.	64.2%	46.6%	69.8%	47.4%	60.4%	43.8%
2-08-2.	65.1%	51.9%	55.8%	42.9%	58.1%	43.9%
2-10-0.	70.4%	42.2%	70.4%	41.3%	70.4%	40.4%
2-10-1.	66.0%	43.2%	69.8%	45.7%	67.9%	46.2%
2-10-2.	69.5%	49.4%	64.4%	44.2%	71.2%	47.7%
2-12-0.	64.8%	47.9%	63.0%	44.2%	68.5%	50.0%
2-12-1.	69.7%	54.1%	60.6%	50.0%	59.1%	50.0%
2-12-2.	69.2%	52.2%	69.2%	52.2%	65.4%	50.7%
2-14-0.	66.7%	43.3%	66.7%	45.6%	66.7%	47.3%
2-14-1.	58.6%	45.9%	66.7%	47.2%	69.0%	48.4%
2-14-2.	66.7%	48.0%	64.8%	47.9%	66.7%	49.3%
2-15-0.	68.8%	41.3%	68.8%	42.3%	56.3%	37.5%
2-15-1.	66.2%	48.4%	67.6%	47.9%	63.2%	49.4%
2-15-2.	70.7%	44.6%	67.2%	43.8%	69.0%	44.9%
2-17-0.	58.6%	51.5%	55.2%	51.6%	58.6%	51.5%
2-17-1.	69.7%	51.7%	59.1%	47.0%	65.2%	51.2%
2-17-2.	70.0%	49.1%	65.0%	48.1%	57.5%	43.4%
2-18-0.	68.8%	46.7%	71.0%	45.5%	69.9%	43.9%
2-18-1.	70.3%	45.9%	65.6%	45.2%	64.1%	43.2%
2-18-2.	63.2%	41.4%	70.1%	42.4%	70.1%	43.3%
2-19-0.	67.2%	42.2%	70.3%	45.0%	62.5%	40.4%
2-19-1.	74.7%	41.6%	65.7%	36.3%	62.6%	35.6%
2-19-2.	66.2%	49.5%	72.1%	49.5%	70.6%	49.0%
2-20-0.	71.0%	49.5%	68.1%	47.0%	66.7%	45.5%
2-20-1.	72.6%	48.4%	66.1%	46.6%	62.9%	47.0%
2-20-2.	67.3%	50.0%	65.4%	46.6%	59.6%	42.5%
2-23-0.	69.6%	45.9%	66.1%	44.6%	62.5%	46.7%
2-23-1.	68.1%	50.0%	61.1%	44.9%	63.9%	43.8%
2-23-2.	68.3%	52.4%	66.7%	47.7%	65.1%	50.0%
AVE.	67.1%	47.7%	65.7%	46.2%	64.1%	46.0%

表 6.3.4　テキストタイリングアルゴリズムの手法を用いた結果 4

36 対話の結果においても、14 対話の時と似たような結果を得た。

ここで、実際のデータをもとに分析を行なった。評価実験においては、候補セグメントとなる数が非常に多いので、候補セグメントとならない箇所を分析してみたところ、特徴的なことは観察されなかった。しかし、「2 時でいいでしょうか？」、「はい」、「それではそうしましょう」、「はいそれでいいです」というような応答詞が多く含まれるやりとりが続くため、類似度が極端に高くなったり低くなったりすることが見られないということが推測できる。

次に、隣接ブロックの新出単語の頻度を用いた手法の境界算出の定数が 1.0 の時の 14 対話の結果を表 6.3.5 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	精度
99atr01.	82.4%	40.0%	82.4%	41.2%	64.7%	37.9%
99atr02.	80.0%	47.1%	75.0%	42.9%	60.0%	37.5%
99chi0a.	80.0%	16.4%	84.0%	17.4%	88.0%	18.6%
99crl01.	75.0%	17.3%	91.7%	22.9%	75.0%	20.0%
99crl02.	85.0%	28.3%	85.0%	29.8%	70.0%	26.9%
99crl03.	88.9%	34.8%	85.2%	35.4%	63.0%	28.8%
99kyo01.	76.9%	32.3%	61.5%	30.8%	69.2%	36.0%
99kyo0a.	83.3%	31.3%	66.7%	28.6%	66.7%	27.9%
99osa01.	95.8%	26.4%	83.3%	25.6%	70.8%	22.7%
99osa02.	87.9%	27.0%	87.9%	27.1%	82.8%	25.8%
99osa0a.	76.5%	33.3%	70.6%	32.4%	70.6%	35.3%
99tsu0a.	90.0%	28.4%	83.3%	26.6%	80.0%	26.4%
99uec0a.	85.0%	26.2%	95.0%	29.7%	95.0%	29.7%
99was03.	84.2%	47.8%	84.2%	48.5%	84.2%	49.2%
AVE.	83.6%	31.2%	81.1%	31.3%	74.3%	30.2%

表 6.3.5　テキストタイリングアルゴリズムでのセグメンテーションの結果 3

この結果においても、前出の隣接ブロックの索引語頻度からの類似度を用いるも

のと同様に再現率がどのブロック内の発話数の設定においても 80%前後で、精度があまりよくないという結果が導かれた。次に同様の 36 対話の結果を表 6.3.6 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	精度
2-04-0.	76.1%	45.6%	76.1%	44.7%	81.8%	47.4%
2-04-1.	83.8%	49.2%	77.0%	47.9%	78.4%	50.0%
2-04-2.	80.4%	48.4%	82.5%	48.8%	80.4%	48.8%
2-05-0.	92.9%	52.7%	78.6%	49.3%	69.0%	48.3%
2-05-1.	74.0%	47.4%	74.0%	48.1%	78.0%	52.0%
2-05-2.	76.3%	45.3%	71.1%	42.9%	76.3%	46.8%
2-08-0.	84.2%	52.5%	82.0%	50.0%	75.4%	50.0%
2-08-1.	75.5%	44.4%	81.1%	46.2%	79.2%	45.2%
2-08-2.	88.4%	51.4%	76.7%	46.5%	76.7%	47.1%
2-10-0.	84.9%	39.1%	75.5%	36.7%	69.8%	35.6%
2-10-1.	81.1%	43.4%	79.2%	42.0%	77.4%	42.7%
2-10-2.	86.0%	43.0%	80.7%	40.7%	82.5%	41.2%
2-12-0.	83.3%	46.9%	81.5%	47.3%	79.6%	51.2%
2-12-1.	77.3%	51.0%	78.8%	51.5%	78.8%	54.2%
2-12-2.	80.8%	48.8%	80.8%	48.8%	82.7%	50.0%
2-14-0.	82.1%	47.8%	71.8%	43.1%	76.9%	46.9%
2-14-1.	85.1%	50.7%	81.6%	47.7%	81.6%	48.0%
2-14-2.	82.7%	45.7%	80.8%	46.2%	71.2%	44.0%
2-15-0.	81.3%	40.2%	70.8%	35.8%	75.0%	37.1%
2-15-1.	83.8%	51.4%	80.9%	46.2%	82.4%	48.3%
2-15-2.	79.3%	44.2%	77.6%	41.7%	81.0%	43.9%
2-17-0.	79.3%	53.5%	79.3%	54.8%	75.9%	55.0%
2-17-1.	84.8%	52.3%	75.8%	40.9%	81.8%	52.4%
2-17-2.	77.5%	44.9%	70.0%	41.8%	70.0%	43.1%
2-18-0.	84.9%	43.6%	81.8%	43.3%	80.6%	41.0%
2-18-1.	82.8%	43.8%	81.3%	43.7%	79.7%	43.2%
2-18-2.	82.8%	42.9%	82.8%	42.9%	85.1%	44.3%
2-19-0.	84.4%	44.3%	78.1%	42.0%	73.4%	40.2%
2-19-1.	88.9%	40.6%	84.8%	38.9%	83.8%	38.4%
2-19-2.	79.4%	47.0%	80.9%	47.8%	83.8%	48.7%
2-20-0.	82.6%	48.3%	75.4%	44.1%	79.7%	47.8%
2-20-1.	87.1%	50.5%	83.9%	50.0%	79.0%	49.5%

2-20-2.	40.8%	45.4%	78.8%	44.1%	80.8%	44.2%
2-23-0.	85.7%	45.7%	75.0%	42.0%	78.6%	45.8%
2-23-1.	84.7%	47.7%	83.3%	47.6%	80.6%	46.4%
2-23-2.	82.5%	45.2%	82.5%	46.8%	82.5%	47.3%
AVE.	81.3%	46.8%	78.7%	45.1%	78.6%	46.3%

表 6.3.6　テキストタイリングアルゴリズムでのセグメンテーションの結果 4

この結果も 14 対話、また隣接ブロックの索引語頻度を用いるものと同様の結果を示している。

また、隣接ブロックの新出単語の頻度を用いた手法の境界算出の定数が 0.5 の時の 14 対話の結果を表 6.3.7 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	精度
99atr01.	70.6%	42.9%	70.6%	42.9%	47.1%	30.8%
99atr02.	75.0%	48.4%	55.0%	40.7%	60.0%	40.0%
99chi0a.	80.0%	17.9%	76.0%	17.3%	72.0%	18.9%
99crl01.	66.7%	16.7%	75.0%	21.4%	58.3%	17.9%
99crl02.	75.0%	28.3%	70.0%	28.6%	45.0%	21.4%
99crl03.	81.5%	34.9%	74.1%	35.7%	51.9%	26.9%
99kyo01.	53.8%	30.4%	64.2%	28.6%	46.2%	31.6%
99kyo0a.	61.1%	27.5%	55.6%	29.4%	55.6%	27.8%
99osa01.	83.3%	25.3%	62.5%	22.1%	58.3%	23.7%
99osa02.	81.0%	26.7%	62.5%	22.1%	70.7%	25.8%
99osa0a.	64.7%	34.4%	52.9%	32.1%	58.8%	35.7%
99tsu0a.	90.0%	31.8%	80.0%	28.9%	76.7%	28.4%
99uec0a.	85.0%	27.0%	85.0%	28.8%	75.0%	26.8%
99was03.	71.1%	49.1%	76.3%	49.2%	71.1%	50.9%
AVE.	74.2%	31.5%	68.5%	30.6%	60.5%	29.1%

表 6.3.7　テキストタイリングアルゴリズム手法による結果 5

この結果においても、前出の隣接ブロックの索引語頻度からの類似度を用いるものと同様にどのブロック内の発話数の設定においても再現率、精度ともによくないという結果であった。索引語頻度を用いる場合よりも同様の条件下（14 対話で定数が 0.5）と比較すると、再現率が若干下がっていることから候補となるセグメントの数は新出単語を用いたもののほうが少なくなっていることが推測される。次に同様の 36 対話の結果を表 6.3.8 に示す。

	ブロック 2		ブロック 3		ブロック 4	
	再現率	精度	再現率	精度	再現率	
2-04-0.	47.1%	42.9%	60.0%	42.9%	60.0%	30.8%
2-04-1.	60.0%	48.4%	45.9%	40.7%	47.5%	40.0%
2-04-2.	72.0%	17.9%	58.8%	17.3%	57.4%	18.9%
2-05-0.	58.3%	16.7%	48.1%	21.4%	46.2%	17.9%
2-05-1.	45.0%	28.3%	57.6%	28.6%	57.6%	21.4%
2-05-2.	51.9%	34.9%	59.0%	35.7%	59.0%	26.9%
2-08-0.	46.2%	30.4%	58.7%	28.6%	56.5%	31.6%
2-08-1.	55.6%	27.5%	47.8%	29.4%	46.3%	27.8%
2-08-2.	58.3%	25.3%	48.1%	22.1%	46.3%	23.7%
2-10-0.	70.7%	26.7%	62.7%	22.1%	60.8%	25.8%
2-10-1.	58.8%	34.4%	64.6%	32.1%	64.6%	35.7%
2-10-2.	76.7%	31.8%	57.5%	28.9%	56.2%	28.4%
2-12-0.	75.0%	27.0%	63.0%	28.8%	63.0%	26.8%
2-12-1.	71.1%	49.1%	73.0%	49.2%	73.0%	50.9%
2-12-2.	64.7%	48.8%	60.7%	48.8%	58.9%	37.9%
2-14-0.	60.0%	47.8%	75.0%	43.1%	75.0%	37.5%
2-14-1.	80.0%	50.7%	57.1%	47.7%	57.1%	18.6%
2-14-2.	75.0%	45.7%	66.7%	46.2%	66.7%	20.0%
2-15-0.	68.8%	40.2%	50.0%	35.8%	48.0%	26.9%
2-15-1.	63.0%	51.4%	51.4%	46.2%	51.4%	28.8%
2-15-2.	69.2%	44.2%	48.8%	41.7%	48.8%	36.0%
2-17-0.	66.7%	53.5%	46.7%	54.8%	46.7%	27.9%
2-17-1.	70.8%	52.3%	50.9%	40.9%	49.1%	22.7%
2-17-2.	68.9%	44.9%	63.4%	41.8%	61.0%	25.8%
2-18-0.	70.6%	43.6%	53.5%	43.3%	51.2%	35.3%
2-18-1.	79.7%	43.8%	66.7%	43.7%	58.3%	26.4%

2-18-2.	70.8%	42.9%	60.0%	42.9%	55.0%	35.8%
2-19-0.	74.3%	44.3%	60.9%	42.0%	60.9%	46.2%
2-19-1.	77.6%	40.6%	73.3%	38.9%	73.3%	41.7%
2-19-2.	72.3%	47.0%	68.2%	47.8%	63.6%	54.8%
2-20-0.	64.6%	48.3%	49.2%	44.1%	47.5%	40.9%
2-20-1.	70.0%	50.5%	54.5%	50.0%	68.7%	41.8%
2-20-2.	70.8%	45.4%	50.0%	44.1%	32.5%	35.8%
2-23-0.	68.8%	45.7%	64.2%	42.0%	72.3%	46.2%
2-23-1.	77.6%	47.7%	60.6%	47.6%	57.6%	41.7%
2-23-2.	70.0%	45.2%	57.7%	46.8%	57.7%	41.8%
AVE.	66.7%	40.7%	58.2%	39.1%	57.1%	32.7%

表 6.3.8　テキストタイリングアルゴリズムの手法の結果 8

この結果も 14 対話、また隣接ブロックの索引語頻度を用いるものと同様の結果を示している。

ここでも、索引語頻度を用いたものと同様に、実際のデータをもとに分析を行なった。評価実験においては、候補セグメントとなる数が非常に多いので、候補セグメントとならない箇所を分析してみたところ、特徴的なことは観察されなかった。しかし、索引語頻度を用いた場合と同様に応答詞を多く含むやりとりが続くため、類似度が極端に高くなったり低くなったりすることが見られないということが推測できる。

また、今回は対話全体における語の分布や推移を見ることができなかったため、同じタスクを扱っている対話の中での話題の変わる箇所をうまく見つけられなかったことも考えられるため、今後は対話全体に対する語の分布、推移を考慮する必要があると思われる。また、テキストタイリングアルゴリズムでは、セグメントを決定する境界を算出する際の定数が 0.5、または 1.0 とするとあったために、それをそのまま使っていたが、その定数自体も変えて実験を必要があると思われる。

第 7 章
重み付けを用いた手法と評価、考察

本章では、重み付けを用いた手法を評価し、評価に基づく考察を行なう。

7.1 重み付けを用いた手法

前章まで、表層的な手がかりを用いた手法と内容的な手がかりを用いた手法をそれぞれ評価、実験してきた。もし、これらのそれぞれの手法がそれなりに有効であれば、それらを組み合わせてセグメンテーションを行なうことも有効である可能性がある。そこで、表層的な手がかりを用いた手法と内容的な手がかりを用いた手法のすべてを組み合わせて最適な手法の組み合わせを探すために、最大エントロピー法を用いる。

最大エントロピー法とは、与えられた制約のもとで、エントロピーを最大化するようなモデルを推定する方法で、基本的には、与えられた制約を満たすモデルの中で最も一様な分布を持つものを選び出すことである[北 99]。

7.2 最大エントロピーを用いた方法

最大エントロピーを用いた方法では、下の表 7.2 に上げた規則を素性とした。

前後の発話の境界に対して

素性 0）接続詞が後ろの発話に存在する場合にセグメントとする

素性 1）前後の発話の繰り返しを考慮する

素性 2）実質的に意味のない語が前の発話に存在する場合にセグメントとする

素性 3）フィラーが後ろの発話に存在する場合にセグメントとする

素性 4）感動詞が後ろの発話に存在する場合にセグメントとする

素性 5）接続詞が前の発話に存在する場合にセグメントとする

素性 6）実質的に意味のない語が後ろの発話に存在する場合にセグメントとする

素性 7）フィラーが前の発話に存在する場合にセグメントとする

素性 8）感動詞が前の発話に存在する場合にセグメントとする

素性 9）索引語頻度を用いたテキストタイリングアルゴリズムの境界算出の定数が 0.5 の時にセグメントとなるものをセグメントとする

素性 10）新出単語を用いたテキストタイリングアルゴリズムの境界算出の定数が 0.5 の時にセグメントとなるものをセグメントとする

素性 11）索引語頻度を用いたテキストタイリングアルゴリズムの境界算出の定数が 1 の時にセグメントとなるものをセグメントとする

素性 12）新出単語を用いたテキストタイリングアルゴリズムの境界算出の定数

が 1 の時にセグメントとなるものをセグメントとする

図 7.2.1　最大エントロピーの素性

なお、この素性は、第 4 章から第 6 章までの結果を考慮したもので、最大エントリピー法によって効果的な素性を自動的に選択できることから多めに設定した。なお、素性 9 から素性 12 までのテキストタイリングに関するものは、この手法中で一番結果のよかった１ブロックの発話数が 3 の設定のもののみを使うこととした。

なお、最大エントロピーを求める際には、公開されている Perl5 で記述された最大エントロピーのプログラムを用いた。また、データが 36 対話と少ないことから、手法の可能性を探るという意味でクローズドテストによる評価のみをすることとした。さらに、[Reynar 97]では、最大エントロピー法を用いてセグメンテーションを行なう際の簡単な決定規則として、下のような数式を満たす規則を使っている。そこで、本論文でも表 7.2.2 の規則をセグメントの決定規則とすることとする。

$$\frac{\text{（セグメントになりうる確率）}}{\text{（セグメントになりうる確率）＋（セグメントになりえない確率）}} > 0.5$$

図 7.2.2 セグメンテーションの決定規則

7.3 最大エントロピーを用いた手法の結果と考察

最大エントロピーを用いた手法では、まず、表 7.2 で上げた素性 0、素性 1、素性 2 を対話データの前後の発話に対して考慮することを基本の素性として精度と再現率を求めた。図 7.3.1 にその結果を示す。なお、本章での図においては、一番左から対話データのデータ番号、再現率、精度の順で記述されており、図の一番下に全データの平均再現率と精度が記述されている。

2-04-0 RECALL: 0.295454545454545; PRECISION: 0.456140350877193

2-04-1 RECALL: 0.364864864864865; PRECISION: 0.473684210526316

2-04-2 RECALL: 0.257731958762887; PRECISION: 0.480769230769231

2-05-0 RECALL: 0.261904761904762; PRECISION: 0.423076923076923

2-05-1 RECALL: 0.276595744680851; PRECISION: 0.448275862068966

2-05-2 RECALL: 0.315789473684211; PRECISION: 0.461538461538462

2-08-0 RECALL: 0.360655737704918; PRECISION: 0.511627906976744

2-08-1 RECALL: 0.283018867924528; PRECISION: 0.441176470588235

2-08-2 RECALL: 0.325581395348837; PRECISION: 0.518518518518518

2-10-0 RECALL: 0.37037037037037; PRECISION: 0.384615384615385

2-10-1 RECALL: 0.320754716981132; PRECISION: 0.414634146341463

2-10-2 RECALL: 0.355932203389831; PRECISION: 0.5

2-12-0 RECALL: 0.277777777777778; PRECISION: 0.416666666666667

2-12-1 RECALL: 0.287878787878788; PRECISION: 0.5

2-12-2 RECALL: 0.307692307692308; PRECISION: 0.516129032258065

2-14-0 RECALL: 0.307692307692308; PRECISION: 0.461538461538462

2-14-1 RECALL: 0.32183908045977; PRECISION: 0.509090909090909

2-14-2 RECALL: 0.277777777777778; PRECISION: 0.428571428571429

2-15-0 RECALL: 0.3125; PRECISION: 0.394736842105263

2-15-1 RECALL: 0.279411764705882; PRECISION: 0.431818181818182

2-15-2 RECALL: 0.344827586206897; PRECISION: 0.416666666666667

2-17-0 RECALL: 0.275862068965517; PRECISION: 0.421052631578947

2-17-1 RECALL: 0.287878787878788; PRECISION: 0.487179487179487

2-17-2 RECALL: 0.3; PRECISION: 0.428571428571429

2-18-0 RECALL: 0.301075268817204; PRECISION: 0.417910447761194

2-18-1 RECALL: 0.28125; PRECISION: 0.409090909090909

2-18-2 RECALL: 0.367816091954023; PRECISION: 0.533333333333333

2-19-0 RECALL: 0.3125; PRECISION: 0.4

2-19-1 RECALL: 0.363636363636364; PRECISION: 0.4

2-19-2 RECALL: 0.367647058823529; PRECISION: 0.595238095238095

2-20-0 RECALL: 0.318840579710145; PRECISION: 0.431372549019608

2-20-1 RECALL: 0.290322580645161; PRECISION: 0.473684210526316
2-20-2 RECALL: 0.307692307692308; PRECISION: 0.421052631578947
2-23-0 RECALL: 0.285714285714286; PRECISION: 0.432432432432432
2-23-1 RECALL: 0.236111111111111; PRECISION: 0.404761904761905
2-23-2 RECALL: 0.26984126984127; PRECISION: 0.425
RECALL_AV: 0.299249724487918; PRECISION_AV: 0.43972853285637

表 7.3.1 最大エントロピーを用いたセグメンテーションの結果 1

この結果によると、再現率が 30%弱、精度が 40%強というあまりよくない結果を得た。以下、最大エントロピー法を用いて効果的な素性を自動的に選択させた。上記の素性に加えて、素性の数を 1 から 5 まで変化させた。次に、素性の数を 2 つ選択した場合（計 5 個の素性）の結果を下の図 7.3.2 に示す。なおこの際、セグメントになる確率を計算するために、素性 10、11 を利用し、セグメントにならない確率を計算するために、素性 5、10 を利用した。

2-04-0 RECALL: 0.704545454545455; PRECISION: 0.466165413533835
2-04-1 RECALL: 0.932432432432432; PRECISION: 0.570247933884298
2-04-2 RECALL: 0.835051546391753; PRECISION: 0.48502994011976
2-05-0 RECALL: 0.904761904761905; PRECISION: 0.535211267605634
2-05-1 RECALL: 0.723404255319149; PRECISION: 0.47887323943662
2-05-2 RECALL: 0.789473684210526; PRECISION: 0.447761194029851
2-08-0 RECALL: 0.754098360655738; PRECISION: 0.567901234567901
2-08-1 RECALL: 0.735849056603774; PRECISION: 0.513157894736842
2-08-2 RECALL: 0.837209302325581; PRECISION: 0.507042253521127
2-10-0 RECALL: 0.796296296296296; PRECISION: 0.5375
2-10-1 RECALL: 0.867924528301887; PRECISION: 0.638888888888889
2-10-2 RECALL: 0.813559322033898; PRECISION: 0.657534246575342
2-12-0 RECALL: 0.740740740740741; PRECISION: 0.481927710843373

2-12-1 RECALL: 0.712121212121212; PRECISION: 0.540229885057471
2-12-2 RECALL: 0.903846153846154; PRECISION: 0.546511627906977
2-14-0 RECALL: 0.794871794871795; PRECISION: 0.53448275862069
2-14-1 RECALL: 0.827586206896552; PRECISION: 0.585365853658537
2-14-2 RECALL: 0.888888888888889; PRECISION: 0.592592592592593
2-15-0 RECALL: 0.8125; PRECISION: 0.443181818181818
2-15-1 RECALL: 0.882352941176471; PRECISION: 0.545454545454545
2-15-2 RECALL: 0.827586206896552; PRECISION: 0.521739130434783
2-17-0 RECALL: 0.620689655172414; PRECISION: 0.514285714285714
2-17-1 RECALL: 0.848484848484849; PRECISION: 0.513761467889908
2-17-2 RECALL: 0.775; PRECISION: 0.462686567164179
2-18-0 RECALL: 0.870967741935484; PRECISION: 0.47093023255814
2-18-1 RECALL: 0.890625; PRECISION: 0.513513513513513
2-18-2 RECALL: 0.885057471264368; PRECISION: 0.487341772151899
2-19-0 RECALL: 0.859375; PRECISION: 0.466101694915254
2-19-1 RECALL: 0.909090909090909; PRECISION: 0.422535211267606
2-19-2 RECALL: 0.838235294117647; PRECISION: 0.471074380165289
2-20-0 RECALL: 0.811594202898551; PRECISION: 0.571428571428571
2-20-1 RECALL: 0.806451612903226; PRECISION: 0.632911392405063
2-20-2 RECALL: 0.846153846153846; PRECISION: 0.586666666666667
2-23-0 RECALL: 0.821428571428571; PRECISION: 0.505494505494505
2-23-1 RECALL: 0.791666666666667; PRECISION: 0.57
2-23-2 RECALL: 0.857142857142857; PRECISION: 0.580645161290323
RECALL_AV: 0.797758485583139; PRECISION_AV: 0.512599358941825

表 7.3.2　最大エントロピーの結果 2

この結果では、再現率が 80%弱、精度が 50%強という結果であった。これを図 8.3.1 の素性が 3 つの時の結果と比較すると、再現率では約 50%、精度は約 10%よくなった。

次に、素性を 4 個選択した場合（計 7 個の素性）の結果を下の図 7.3.3 に示す。な

おこの際、セグメントになる確率を計算するために、素性 6、8、10、11 を利用し、セグメントにならない確率を計算するために、素性 5、6、10、11 を利用した。

2-04-0 RECALL: 0.75; PRECISION: 0.532258064516129
2-04-1 RECALL: 0.891891891891892; PRECISION: 0.605504587155963
2-04-2 RECALL: 0.824742268041237; PRECISION: 0.559440559440559
2-05-0 RECALL: 0.928571428571429; PRECISION: 0.629032258064516
2-05-1 RECALL: 0.574468085106383; PRECISION: 0.473684210526316
2-05-2 RECALL: 0.842105263157895; PRECISION: 0.524590163934426
2-08-0 RECALL: 0.737704918032787; PRECISION: 0.584415584415584
2-08-1 RECALL: 0.773584905660377; PRECISION: 0.585714285714286
2-08-2 RECALL: 0.813953488372093; PRECISION: 0.555555555555556
2-10-0 RECALL: 0.814814814814815; PRECISION: 0.6875
2-10-1 RECALL: 0.754716981132076; PRECISION: 0.701754385964912
2-10-2 RECALL: 0.745762711864407; PRECISION: 0.771929824561403
2-12-0 RECALL: 0.703703703703704; PRECISION: 0.520547945205479
2-12-1 RECALL: 0.712121212121212; PRECISION: 0.61038961038961
2-12-2 RECALL: 0.903846153846154; PRECISION: 0.652777777777778
2-14-0 RECALL: 0.82051282051282; PRECISION: 0.627450980392157
2-14-1 RECALL: 0.839080459770115; PRECISION: 0.682242990654206
2-14-2 RECALL: 0.851851851851852; PRECISION: 0.666666666666667
2-15-0 RECALL: 0.75; PRECISION: 0.48
2-15-1 RECALL: 0.867647058823529; PRECISION: 0.621052631578947
2-15-2 RECALL: 0.810344827586207; PRECISION: 0.55952380952381
2-17-0 RECALL: 0.620689655172414; PRECISION: 0.529411764705882
2-17-1 RECALL: 0.878787878787879; PRECISION: 0.597938144329897
2-17-2 RECALL: 0.825; PRECISION: 0.559322033898305
2-18-0 RECALL: 0.817204301075269; PRECISION: 0.567164179104478
2-18-1 RECALL: 0.875; PRECISION: 0.595744680851064
2-18-2 RECALL: 0.873563218390805; PRECISION: 0.575757575757576

2-19-0 RECALL: 0.859375; PRECISION: 0.578947368421053
2-19-1 RECALL: 0.898989898989899; PRECISION: 0.511494252873563
2-19-2 RECALL: 0.838235294117647; PRECISION: 0.59375
2-20-0 RECALL: 0.840579710144927; PRECISION: 0.707317073170732
2-20-1 RECALL: 0.82258064516129; PRECISION: 0.761194029850746
2-20-2 RECALL: 0.846153846153846; PRECISION: 0.745762711864407
2-23-0 RECALL: 0.803571428571429; PRECISION: 0.625
2-23-1 RECALL: 0.791666666666667; PRECISION: 0.678571428571429
2-23-2 RECALL: 0.857142857142857; PRECISION: 0.72972972972973
RECALL_AV: 0.78810716879016; PRECISION_AV: 0.594300996355869

表 7.3.3 最大エントロピーの結果 3

この結果では、再現率が 80%弱、精度が 60%弱という結果を得た。これを図 8.3.2 の時の素性が 5 つの時と比較すると、再現率はほとんど変わらないが、精度では約 10%よい結果が得られた。

この他に、素性をそれぞれ 8 つまで増やして実験を行ない、図 8.3.2 との比較を行ない、再現率、精度を合わせて考えて検討した結果、この場合よりもよい結果を得ることができなかった。これにより、表層的手がかりに内容的手がかりを加えた手法でセグメンテーションを行なうことはある程度有効であることが推測され、また、素性を増やせば結果がよくなるということはなく、素性の選び方が重要であるということがわかる。

第 8 章
おわりに

本研究では、メールやチャットといった新しいコミュニケーション形態での情報をうまく活用する技術を確立することを念頭におきながら、従来のコミュニケーション形態のもっとも一般的な対話に関して、話題のまとまりを自動的に抜き出す方法を文書対象の既存の研究から検討し、評価実験を行なった。

分析に利用した対話データは、数は 14 対話と 36 対話の 50 対話と数は多くはないが、課題のバリエーションがあり、付与情報が豊富なデータである。そういった付与情報をいろいろ使える手法として、接続標識、談話標識、フィラーに代表される手がかり語を利用したり、品詞情報を利用したり、内容的には単独では意味を持た

ないと判断される実質的に意味のない語といったような表層的手がかりと、発話で発せられる単語の分布に着目し、文書を対象としたテキストタイリングアルゴリズムを用いて内容的手がかりとの両面から評価実験を行なってきた。

表層的手がかりに関しては、機能を重視した手がかり語のうちの談話標識では、14 対話のみの対象であるが、85.1%と比較的高い精度でセグメンテーションを行なうことが可能であった。また、形態素解析の結果の手がかり語のうちの接続詞に関しては、14 対話を対象にした場合、72%、36 対話を対象にした場合、71.7%と比較的高い精度でセグメンテーションを行なうことが可能であった。実質的に意味のない語を用いたセグメンテーションにおいて、14 対話を対象に人手により作成した実質的に意味のない語のリストを使い、そのリスト中の語を話題のまとまりの終わりであると見なした場合、14 対話では 82.8%と比較的高い精度でセグメンテーションを行なうことができ、36 対話を対象にした場合でも、57.1%の精度でセグメンテーションを行なうことができた。機能を重視した手がかり語である談話標識に関しては、人手によるタグ付けが要求されるのが現状であり、実質的に意味のない語に関しては明確な定義がなされていないために、人手によるタグづけもできないという問題点がある。

また、内容的手がかりを用いた手法に関しては、文書を対象にしたテキストタイリングアルゴリズムを対話に適応した結果、索引語を用いたセグメンテーションにおいても、新出単語を用いたセグメンテーションにおいても、30%～50%弱とあまりよい結果を得ることができなかった。これは、今回対対話において、出現する単語の数、種類とともに統計的処理が有効に働くために十分ないことが言える。また、対話に対してブロックの大きさを単位でブロックを決めることができなかったことが影響している可能性がある。

さらに、表層的手がかりと内容的手がかりの両方を考慮するために、重み付けを行なう手法として、最大エントロピー法を用いて評価を行なった。この結果、表層的な手がかりだけを素性として使った場合は、44%の精度しか得られなかったが、内容的な手がかりを素性に加えることにより、59.4%の精度を得ることができた。ここから、表層的な手がかりと内容的な手がかりを組み合わせることはある程度有効であるということが言える。また、素性数に関しても、素性数 7 の時が最もよい精度を得ることができたが、それ以上、それ以下では素性数 7 の精度を越えることは

なかった。ここから、有効であると思われる素性をいくつも組み合わせればいいというものではなく、最適な組み合わせというものを考慮する必要性があると考えられる。

また、今回のすべての手法を評価する際に使用した対話データ自体もタグ付けが完全ではなく、セグメンテーションの評価に使用した正解セグメントタグ自体にも問題がある可能性がある。正解セグメントに関しては、対話のやりとり構造を考慮したタグ方式[山下 99]をもとに、内容の変化度合いで 2 段階に分けたセグメントタグの付与が行なわれているが、やりとり構造に重要な役割を果たす働きかけ発話の定義が十分ではないなどの理由から誰がセグメントタグを付与しても一致するというようなことが行なえないのが現状である。

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。心から感謝のいたします。

また、普段の研究活動において、石崎研究室の学生の皆様、また、知識創造論講座の皆様方には、多くのヒント、アドバイスを頂き、心より感謝いたします。特に、松永政幸さん、須藤由加さんには実験に使用するデータの修正等でご協力していただきました。さらに、今回実験に使用したデータは、NTT サイバースペース研究所の許可を得て利用させていただきました。ここに感謝の意を込めて明記させていただきます。

参　考　文　献

[Grosz 86]　Babara J.Grosz　Candace L.Sinder：Attention,Intentions,and the Structure of Discourse，Computational Linguistics Vol.12 175-204 1986

[中里 99]　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.5 900-906 1999

[人工知能学会談話対話におけるコーパス利用研究グループ 99]　人工知能学会　談話・対話におけるコーパス利用研究グループ：日本語スラッシュ単位（発話単位）ラベリングマニュアル Ver1.0，人工知能学会 1999

[金 00]　金美慶：対話セグメンテーション方式に関する研究，北陸先端科学技術大

学院大学修士論文 2000
[Hearst 94] Marti A. Hearst：Multi-Paragraph Segmentation Of Expository Text，ACL94 1-8
[Hearst 97] Marti A. Hearst：TextTiling:Segmenting Text into Multi-paragraph Subtopic Passage，Computational Linguistics Vol.23 No.1 1997
[望月 99] 望月源 本田岳夫 奥村学：複数の表層的手がかりを統合したテキストセグメンテーション，自然言語処理 Vol.6 No.3 1999
[山下 99] 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の方式の検討，人工知能学会誌 Vol.14 No.2 282-289 1999
[北 99] 北研二：言語と計算 4 確率的言語モデル，東京大学出版会 1999
[Reynar 97] Jeffrey C. Reynar Adwait Ratnaparkhi：A Maximum Entropy Approach to Identify Sentence Boundaries，ANLP97 1997