センタリング理論と対象知識に基づく談話構造解析システム DIA

梅澤 俊之[†]・原田 実^{††}

自動要約，照応解析，質問応答，評判分析などの応用では，文章中の文間の役割的關係や話題の推移の解析が必要になりつつある．本研究ではセンタリング理論と対象知識に基づき談話中の話題の移り変わりを意味的に解析し談話構造を生成する手法を提案し，これに基づく談話構造解析システム DIA を開発した．本研究ではセンタリング理論を拡張し文の依存先の決定に用いる．また，語の語意が表す概念の部分／属性関係，上位一下位関係，類似関係といった対象に関する意味的關係を EDR 電子辞書中の共起辞書と概念体系辞書から抽出し，文間の接続関係の決定に用いる．談話の中で話題の推移を表すために談話構造木を定義した．談話構造木では，句点で区切られた文をノードとし，各ノードはただ1つの親ノードを持つ．また，文を接続しているアークに9種類の文間接続関係（詳細化，展開，原因—結果，逆接，遷移，転換，並列，例提示，質問—応答）を付与する．談話構造を決定する手法としては，拡張したセンタリング理論を元に各文に対してそれより前にある文に対して親ノードとなる可能性を得点化し，最高点のものを親ノードとする．次に各文ノード間のアークに対して，その接続関係が9種のそれぞれである可能性を評価する経験的なルールを36個定め，最高得点を得た関係ラベルをアークに付与する．評価実験の結果，接続先の特定では82%，文間接続関係の判定では81%の精度を実現した．

キーワード：談話構造解析，センタリング理論，文間接続関係，対象知識，文脈解析

Discourse Structure Analysis System DIA Based on Centering Theory and Object Knowledge

TOSHIYUKI UMEZAWA[†] and MINORU HARADA^{††}

We propose the method of the discourse structure analysis based on the centering theory and the domain knowledge. We enhanced the centering theory to utilize it for the analysis of dependency relation among sentences. Moreover, object knowledge among objects such as entire-part relation, parent-child relation, object-attribute relation, resemblance are extracted from the co-occurrence dictionary and concept dictionary in EDR electronic dictionary, and are used for the decision of the functional relation between sentences. We defined the discourse structure tree to show the transition of the topic in the discourse. Here, the sentence punctuated by the period makes a node, and each sentence node has only one parent node. Moreover, nine kinds of connection relationships between sentences: DETAIL, EXPLICATE,

[†] 青山学院大学大学院理工学研究科知能情報コース, Intelligent Information Course, Graduate School of Science & Engineering, Aoyama Gakuin University

^{††} 青山学院大学理工学部情報テクノロジー学科, Dept. of Integrated Information Technology, Faculty of Science & Engineering, Aoyama Gakuin University

CONTRAST, CAUSE-RESULT, TRANSITION, CONVERSION, PARALLEL, EXAMPLE and QUESTION-RESPONSE are given to each arc between sentence nodes. Next, we have developed 36 rules to evaluate the possibility of the relation between sentences being each one of the nine kinds of relation types. These rules are applied to each arc, and the relation label which obtained the highest score is given to the arc. The evaluation experiment shows the accuracy of 82% and 81% respectively in deciding the connection destination and the relationship label.

Key Words: *Discourse structure analysis, Centering theory, Inter-sentence relationship, Object knowledge, Context analysis*

1 はじめに

自然言語処理の研究分野において、1文を対象にした研究は盛んに行われてきた。特に、形態素解析や構文解析は実用レベルに達しており、様々な自然言語を対象とした応用研究において、基礎処理として使用されている。しかし、高度な文章処理を目的としている応用研究、例えば文章要約や照応解析、質問応答、評判分析などは、当然ながら1文を対象にしているわけではなく、高い精度を実現するためには、文章中の話題のまとまりや文間の接続関係といった談話構造の理解が必要になる。このような談話構造解析を用いれば、文章要約(田中, 面来, 野口, 矢後, 韓, 原田 2006)では話題のまとまりを考慮した自然な要約が可能になり、照応解析(南, 原田 2002)では先行詞候補を探索する範囲を談話構造木の照応詞と根を結ぶ経路上へと高い精度で絞り込むことができ、質問応答システム(加藤, 古川, 蒲生, 韓, 原田 2005)では理由や原因の回答抽出が容易になることが期待される。

談話構造解析の従来研究では様々なモデルが提案されてきた。何を基本単位とするか、単位間の関係、談話構造のモデルなど研究者により様々である。談話構造のモデルとしては文を基本単位とした木構造モデルが一般的である。黒橋ら(黒橋, 長尾 1994)は文間に11種類の結束関係(並列, 対比, 主題連鎖, 焦点—主題連鎖, 詳細化, 理由, 原因—結果, 変化, 例提示, 例説明, 質問—応答)を定義し、手掛かり表現・主題連鎖・文間の類似性に着目し判定している。横山ら(横山, 難波, 奥村 2003)は8種類(因果, 背景, 呼応, 並列, 対比, 転換, 補足, 例示)の係り受け関係をSVMを用いた機械学習により判定している。Marcu(Marcu 2002)は木構造モデルではなく、連続する2文に限り、4種類の接続関係(CONTRAST, CAUSE-EXPLANATION-EVIDENCE, CONDITION, ELABORATION)を大量のテキストデータを用いた用例利用型の手法で判定している。山本ら(山本, 齊藤 2008)は、同様の手法で、6種類の接続関係(累加, 逆接, 因果, 並列, 転換, 例示)を判定している。

以上のように談話構造解析の従来研究では様々な解析方法が提案されているが、大きく2つ

の問題がある。1 つ目は、文の話題の中心である焦点の推移を詳細に分析できていないという問題である。焦点はその文を象徴する最も重要な手掛かりであり、談話構造解析には欠かせない要素である。2 つ目は、基本的に接続詞や文末表現、同一語の出現など、表層的な情報に基づいているという問題である。特に、接続詞が文中に現れる頻度はあまり高くない。シソーラスを用いて類義情報を取り入れている研究もあるが、そもそも利用されている意味解析の精度が低く類義判定が信頼性に欠けること、また談話では主題の属性や部分などへの話題の変化が多く見られ、類義情報のみでは文間のつながりを適切に把握できないなどの問題点がある。

本研究では精度の高い談話構造解析を実現するため、談話の結束性を評価するセンタリング理論を談話構造解析に導入することで、談話の焦点の推移を詳細に捉えることを可能にする。そして部分／属性関係など 2 語が表す概念間の意味的關係を定めるにあたって、原田らが開発した意味解析システム Sage (語意精度 95%, 深層格精度 90%) (原田, 尾見, 岩田志, 水野 1999; 原田, 水野 2001; 原田, 田淵, 大野 2002) を用いて各語の意味 (概念) を高精度に定め、さらに EDR 電子辞書 (1995) から抽出した概念間の部分／属性関係を対象知識として、話題の部分／属性への展開などの検出に用いる手法を提案する。

2 提案手法

本研究では談話構造を表すモデルとして談話の話題の推移を表すために「談話構造木」という木構造を定義する。談話構造木では、句点で区切られた文をノードとし、各ノード文はただ 1 つの親ノードを持つ。そして文ノードを接続しているアークに 9 種類の文間接続関係 (詳細化, 展開, 原因—結果, 逆接, 遷移, 転換, 並列, 例提示, 質問—応答) を付与する。以下に例を示す。

N 自動車は 19 日、新車発表会を開催。N 自動車としては初となる電気自動車を公開した。電気自動車は、走行時に CO₂ を排出しないとして注目を集めている。しかし、一充電あたりの航続距離が短いなどの問題点もある。販売を伸ばしているハイブリット車への対抗として、N 自動車は巻き返しを図る考えだ。

図 1 の例文では、1 文目と 2 文目ではともに「N 自動車」を主な話題としている。そして 3 文目では 2 文目に登場した「電気自動車」へと話題が展開している。4 文目では、3 文目の話題「電気自動車」についてさらに情報を付加している。そして 5 文目ではまた「N 自動車」へと話題が戻っている。例文では、大まかに分けると「N 自動車」とそこから派生した「電気自動車」についての話題が存在し、図 1 の談話構造木でその話題の推移を表すことができる。

本研究ではセンタリング理論 (Grosz, Joshi and Weinstein 1995) と対象知識に基づき談話中

図 1 談話構造木の例 1

の話題の移り変わりに着目した談話構造解析の手法を提案する。センタリング理論とは文の話題の中心である焦点の推移に着目して文間の結束性をモデル化した理論である（第4章）。また、対象知識とは、語の語意が表す概念の部分／属性関係、上位一下位関係、類義関係といった2概念間の意味的關係をEDR電子辞書の共起辞書と概念辞書から抽出したものである（第3章）。本研究の手法では対象知識により話題の部分や属性などへの展開といった概念の意味的關係を考慮してセンタリング理論を拡張し、文の話題の推移を的確に捉えられる談話構造解析を可能にする。その結果として、先に述べたように、文をノードとする談話構造木を機械的に生成する。この時、焦点の推移に基づく文間の結束性を文を表すノード間のアークとして表現する。さらにアークにこの文間の焦点の推移が何を意味するのかを表す理由や展開といった9種類の文間接続関係ラベルを割り振ることで、文同士の役割的關係を明らかにする。これによって、質問応答や照応解析の解の探索範囲を絞り込んだり、自動要約で主たる話の流れを示す評価基準を得られることが期待できる。

なお、Groszらのセンタリング理論では代名詞の扱いに関する規則が与えられているが、現在筆者らの環境ではゼロ代名詞や指示代名詞の先行詞を高い精度で特定する技術を確立できていないので、それらを考慮するとかえって談話構造木の構築精度を下げる可能性があるため、本研究では代名詞を扱わないことにした。

談話構造木を以下のプロセスを経て構築する。

- (1) 形態素・構文・意味解析
- (2) 談話構造木の構築
- (3) 文間接続關係の判定

まず、形態素・構文解析を Juman・Knp (黒橋, 長尾 1998) を用いて、意味解析を Sage を用いて行う。これにより談話に含まれる各文は、形態素、文節に分割され、それぞれに EDR 辞書中の品詞と語意 (EDR 辞書で定義された約 40 万概念のどれか) が付与され、文節間の係り受け関係には役割関係を表す深層格 (EDR 辞書で定義されたものに Sage で追加された 30 種のどれか) が付与される。つぎに談話構造木の構築と文間接続関係の判定を分割して行う。談話構造木の構築は、センタリング理論、表層パターン、文間距離に着目した手法を用いて行う (第 5 章)。センタリング理論は Grosz (Grosz, Joshi and Weinstein 1995) のものを、本研究で提案する対象知識を用いて拡張して使用する。最後に、構築された談話構造木中のリンクに文間接続関係を付与する (第 6 章)。文間接続関係の判定には、接続詞、主題、モダリティ・テンス・アスペクト、語意で表される概念間の部分／属性関係などの対象知識に着目したルールを用いて判定する。

本研究の提案する手法では、センタリング理論と対象知識で話題の推移を的確に捉え、談話構造木の構築と文間接続関係の判定のプロセスを分割して行うことにより、文間接続関係の判定を単純化することが可能になる。なお、研究を具体的に進めるにあたって、文間接続関係の定義やその判定ルールの作成においては、対象となる談話として、新聞や Web の報道・解説・論説記事を用いたので、研究成果は精度の信頼性においてこれらの談話が対象になるが、談話構造構築方式はより一般的に適用できると考えている。ただし、会話文などのように言外の指示表現が多用される分野では精度が落ちることが想定される。

本稿では、まず提案手法の基本となる対象知識について 3 章で、センタリング理論について 4 章で述べた後、実際の談話構造解析のプロセスについて、5 章で談話構造木の構築方法、6 章で文間接続関係の判定方法について述べる。

3 対象知識

談話の中でも、特に論文や新聞などの報道・解説・論説といった文章では、まず始めにその談話が何について述べられているものなのかを表す談話全体の話題 (大話題) が示され、続いてその大話題を説明するために、大話題に関連した幾つかのさらに詳細な話題 (小話題) が述べられるといった話題の階層構造を取る場合が多い。小話題としては大話題の部分／属性、下位、類義などが現れやすい。以下を例文として説明する。

相模原市は、神奈川県北部にある都市。人口は 70 万人を超え、神奈川県内では横浜市、川崎市について第 3 位の人口規模を擁する。特に 20 代、30 代、50 代周辺の人口が多く、市全体を活気ある雰囲気になっている。市内に大学が多いことで、学生の街としての顔も併せ持つ。

図 2 対象知識の判定方法

例文では、まず冒頭で談話全体の話題である大話題「相模原市」が示されている。そして、2 文目、3 文目で大話題「相模原市」(Sage で求めた語意は 0f2ff2: 相模原市という市)の部分／属性「人口」(3c0fa3: 一定の地域に住む人の数)が、4 文目では部分／属性「大学」(1e8598: 高等教育の中核となる学術研究および教育の最高機関)が小話題として説明されている。

したがって、例文のような文章を談話解析するためには、「人口」が「相模原市」の部分／属性であるという対象一部分／属性関係といった知識が必要になる。本研究ではこのような話題となる対象間の関係を対象知識と呼ぶ。対象知識としては、対象一部分／属性関係のほかに、上位一下位関係、類義関係を考える。

概念間に意味的關係があるかどうかの判定(図 2)では、2 つの語の語意が表す 2 概念をシソーラスと共起辞書、品詞、概念類似度を用いた対象知識判定規則に照らし合わせ、適合した場合に当該 2 概念間に対象一部分／属性関係、上位一下位関係、類義関係があると考え。シソーラスには EDR 概念辞書、共起辞書には EDR 日本語共起辞書((株)日本語電子辞書研究所 1995) [9] を用いる。以下で対象知識のそれぞれの、対象知識判定規則について述べる。

3.1 対象一部分／属性関係判定規則

対象一部分／属性関係とは表 1 のような概念間の関係を示す。ここで注意したいのは、本研究では、ある概念の部分概念と属性概念を明確に分類しないということである。本研究では、あくまで小話題として現れやすい概念を抽出することを目的としている。

対象一部分／属性関係の判定規則には、主に共起辞書を用いるものとシソーラスでの上位概念ペアを用いるものの 2 種類がある。

共起辞書を用いるものは、共起辞書中の共起関係子が助詞「の」であるレコードに着目する。その理由は、「自動車のエンジン」というように、対象と部分／属性を直接結びつける代表的な助詞として「の」が用いられるからである。対象とするレコードの検索には対象概念と部分／

表 1 対象一部分／属性関係の例

対象概念	部分／属性概念
自動車	エンジン
人	年齢
PC	メモリ

表 2 共起辞書レコードの検索

係り側	共起関係子	受け側	適合条件
対象概念	の	部分／属性概念	このような共起レコードがある
対象概念	の	概念 X	このような共起レコード中の概念 X と部分／属性概念が類義
概念 Y	の	部分／属性概念	このような共起レコード中の概念 Y と対象概念が類義

属性概念のペアを対象とし、完全一致だけでなく類義のレコードも対象とすることで対象一部分／属性概念の可能性のあるものを漏れなく抽出する。レコードの検索方法は表 2 の 3 種類を用い、適合条件に一致すれば該当レコードありとする。

ここで類義かどうかの判定には、シソーラス上の距離に基づく以下の式 (加藤, 古川, 蒲生, 韓, 原田 2005) を用いて類似度を計算し、その類似度が閾値 (本報告では 0.85 とした) 以上のものを類義とする。

$$\text{類似度 (a, b)} = \frac{2(1-r)dc}{2(1-r)dc + (1-r^{da}) + (1-r^{db})} \quad (1)$$

r : 公比 ($r \neq 1.0$)

dc : 共通上位概念までの概念の深さ

da : 共通上位概念から語 a までの深さ

db : 共通上位概念から語 b までの深さ

ただし、助詞「の」の用法には、対象が部分／属性を修飾する用法のほかに、「彼の走る姿」といった主格の用法や、「黄色の花」といった対象をその状態が修飾するもの、「13 時の会議」といった時間による修飾などの用法がある。これらの用法を取り除くため、該当レコードに対して、深層格や品詞、シソーラスでの上位概念による以下の規則を用いてさらに絞り込みを行う。

- (1) 対象とする深層格は modifier (修飾), element-of (要素), part-of (部分)
- (2) 品詞は数詞を除く名詞
- (3) 対象候補は上位概念に「時」, 「状態」, 「物やものに対する指示的な呼称」, 「方向」, 「複数のものの関係によって決まる位置」, 「部分」を持たない

以上の共起辞書レコードの検索，深層格や品詞，シソーラスでの上位概念による規則に適合したものを対象一部分／属性関係とする。

つぎに，上位概念ペアを用いた対象一部分／属性関係の抽出について述べる。EDR 概念辞書には「人間の属性」や「動物の部分」，「具体物の属性」，「機械の部品」などの概念が含まれる。部分／属性概念候補がこれらを上位概念に持ち，さらに対象概念候補がそれら属性を持つにふさわしい概念（表 3 の左列に列挙した概念）を上位概念に持てば，対象一部分／属性関係ありとする。例えば部分／属性候補が「人間の属性」を上位概念に持ち，対象概念候補が「人間」を上位概念に持てば，対象一部分／属性関係ありとする。これらの対象一部分／属性関係を持つ上位概念のペアの一覧を表 3 に示す。

3.2 上位一下位関係判定規則

上位一下位関係とは表 4 のような概念間の関係を指す。上位一下位関係はシソーラスを用いることで容易に抽出することができる。ただし，シソーラス中のすべての上位一下位関係を持つ概念を対象にしてしまうと意味が離れすぎてしまうため，2 概念間に適当な概念間距離閾値（本報告では，距離 2）を設定する。

表 3 部分／属性関係を持つ概念の上位概念ペア一覧

対象概念の上位概念	部分／属性概念の上位概念
人間	人間の属性
人間	動物の部分
人間	生物の生理的属性
動物	動物の部分
動物	生物の生理的属性
具体物	具体物の属性
機械	機械の部品
自動物	機械の部品
植物	植物の部分
もの	物事の構造や内容
もの	全体部分関係，量などで限定した静物や擬似静物や人間
催し	年や月や日でとらえた時
催し	場所
人間または人間と似た振る舞いをする主体	呼び名
もの	呼び名
飲食物を供する店	料理名
世界のあらゆる場所	地名で示す地域
地球上の全ての地域	地名で示す地域

表 4 上位一下位関係の例

上位概念	下位概念
哺乳類	猫
植物	ひまわり
家電	冷蔵庫

3.3 類義判定規則

類義に関してもシソーラスを用いることで容易に抽出することができる. 式 (1) に基づき類似度を計算し, 適当な閾値 (本報告では 0.85) 以上の 2 概念間に類義関係ありと判断する.

4 センタリング理論

センタリング理論 (Grosz, Joshi and Weinstein 1995) とは, 文の焦点の移り変わりに着目して談話の結束性をモデル化したものである. センタリング理論では, 談話単位中の各文 Si^1 に $Cf(Si)$ (Forward-looking centers) を, また各文 Si とその前に出現する各文 Sj に $Cb(Si, Sj)$ (Backward-looking center) を定義する. $Cf(Si)$ は文 Si に出現する要素 (話題) のリストであり, 優先順位によりソートされている. $Cb(Si, Sj)$ は, ただ一つの要素を持ち, 文 Sj から文 Si に談話が推移した時の焦点をあらわしている. 本研究では, 以下のように定義する.

$Cf(Si)$: Forward-looking centers

- 文 Si に出現する名詞節と主辞がサ変名詞である動詞節を要素とするリスト
- 以下の優先順位により降順にソートされている
主題 (ハ格) > ガ格 > ニ格 > ヲ格 > その他

$Cb(Si, Sj)$: Backward-looking center

- $Cf(Sj)$ の要素で Si に含まれる名詞節と, 同一または同義または対象知識 (第 3 章) で判定される部分/属性や上位/下位関係のある概念を表す語を主辞とする文節のうちソート順で最上位要素
- ルートノード文では $Cf(Si)$ の最上位要素
- $i > j$

文の 3 つ組 Si, Sj, Sk ($i > j > k$) に対して, Cb の推移 $Cb(Sj, Sk) \rightarrow Cb(Si, Sj)$ の値を表 5 ² の

¹ Grosz 1995 では発話 U としているが本稿では処理対象単位が文であるので S とした.

² Grosz 1995 では CONTINUATION, RETAINING, SHIFTING の 3 種類.

表 5 TRANSITION の分類

CONTINUATION	$Cb(S_i, S_j) = Cb(S_j, S_k)$ であり $Cf(S_i)$ の最上位要素
RETAINING	$Cb(S_i, S_j) = Cb(S_j, S_k)$ であり $Cf(S_i)$ の最上位要素以外
SHIFTING	$Cb(S_i, S_j) \neq Cb(S_j, S_k)$
NOTHING	$Cb(S_i, S_j) = \phi$

ように定義する．これを TRANSITION と呼ぶ．談話構造木の構築で TRANSITION $Cb(S_j, S_k) \rightarrow Cb(S_i, S_j)$ を求める際には S_k は確定しており， S_k は S_j の親ノードである．

TRANSITION は焦点の連続性を評価している．そして，この TRANSITION の種類により文間の結束性の強さが示される．結束性の強さは以下の順に従う．

CONTINUATION > RETAINING > SHIFTING > NOTHING

以下の例文でセンタリング理論における Cf 、 Cb と TRANSITION の判定例を示す．

例文 a

幸子は夕飯の材料が足りないことに気づいた。そこで幸子は弟に買い物を頼んだ。しかし、弟は幸子に嫌だと言った。弟はゲームに夢中だった。

- a1. 幸子は夕飯の材料が足りないことに気づいた。
 $Cf(Sa1)$: [幸子, 材料, 夕飯]
 $Cb(Sa1, \phi)$: [幸子]
- a2. そこで幸子は弟に買い物を頼んだ。
 $Cf(Sa2)$: [幸子, 弟, 買い物]
 $Cb(Sa2, Sa1)$: [幸子]
TRANSITION $Cb(Sa1, \phi) \rightarrow Cb(Sa2, Sa1)$: CONTINUATION
- a3. しかし、弟は幸子に嫌だと言った。
 $Cf(Sa3)$: [弟, 幸子]
 $Cb(Sa3, Sa1)$: [幸子]
TRANSITION $Cb(Sa1, \phi) \rightarrow Cb(Sa3, Sa1)$: RETAINING
 $Cb(Sa3, Sa2)$: [幸子]
TRANSITION $Cb(Sa2, Sa1) \rightarrow Cb(Sa3, Sa2)$: RETAINING
- a4. 弟はゲームに夢中だった。

Cf(Sa4) : [弟, ゲーム]

Cb(Sa4, Sa1) : [ϕ]

TRANSITION Cb(Sa1, ϕ) $\rightarrow$ Cb(Sa4, Sa1) : NOTHING

Cb(Sa4, Sa2) : [弟]

TRANSITION Cb(Sa2, Sa1) $\rightarrow$ Cb(Sa4, Sa2) : SHIFTING

Cb(Sa4, Sa3) : [弟]

TRANSITION Cb(Sa3, Sa2) $\rightarrow$ Cb(Sa4, Sa3) : SHIFTING

例文 a では, 文 Sa4 で比較先の文 Sj により TRANSITION が分かれている. Cb(Sa4, Sa2) と Cb(Sa4, Sa3) のときは TRANSITION が SHIFTING となっているが, Cb(Sa4, Sa1) の場合は NOTHING である. これはつまり, 文 Sa4 は, 文 Sa1 よりも文 Sa2 や Sa3 との間の結束性が高いことを表している.

次の例では, 部分／属性で Cb が決定されている.

例文 b

N 社は新型自動車を発表した。新型自動車は優れた環境性能を実現。燃費は、23.0 km/l とクラストップレベル。

b1. N 社は新型自動車を発表した。

Cf(Sb1) : [N 社, 新型自動車, 発表]

Cb(Sb1, ϕ) : [N 社]

b2. 新型自動車は優れた環境性能を実現。

Cf(Sb2) : [新型自動車, 環境性能, 実現]

Cb(Sb2, Sb1) : [新型自動車]

TRANSITION Cb(Sb1, ϕ) $\rightarrow$ Cb(Sb2, Sb1) : SHIFTING

b3. 燃費は、23.0 km/l とクラストップレベル。

Cf(Sb3) : [燃費, 23.0 km/l, クラストップレベル]

Cb(Sb3, Sb1) : [燃費 (新型自動車)]

TRANSITION Cb(Sb1, ϕ) $\rightarrow$ Cb(Sb3, Sb1) : SHIFTING

Cb(Sb3, Sb2) : [燃費 (新型自動車)]

TRANSITION Cb(Sb2, Sb1) $\rightarrow$ Cb(Sb3, Sb2) : CONTINUATION

例文 b では, 文 Sb3 の Cb が先行文内の「自動車」の部分／属性概念である「燃費」である. ま

表 6 Cb 決定タイプ

実現タイプ	説明
EQUAL	同一概念または同義概念
ATTRIBUTE	部分／属性概念
LOWER	下位概念
SIMILAR	下位以外の類義概念

た、文 Sb3 で比較先の文 Sj が Sb1 の時の Cb(Sb3, Sb1) の場合は SHIFTING, Sb2 の時の Cb(Sb3, Sb2) の場合は CONTINUATION と分かれており、文 Sb3 は文 Sb2 との結束性が高い。

Cb(Si, Sj) が部分／属性概念などで決定される場合と同一概念や同義概念で決定される場合を談話構造木構築の際に区別するため、Cb 決定タイプを表 6 のように設定する。

Cf(Si) の要素と同一概念または同義概念が含まれない場合は、EQUAL 以外の決定タイプが採用される。この場合の優先順位は、ATTRIBUTE, LOWER, SIMILAR の順である。

第 5 章で詳しく述べるが、本研究では、談話構造木の構築にこの TRANSITION の種類における結束性の強さを主な指標として用いる。談話構造木は談話における話題の推移を表す木構造であるので、同じ焦点 Cb を持つ文同士が接続されることが望ましい。このことは、結束性の強い TRANSITION (CONTINUATION や RETAINING) になる文同士を接続したリンクを多く含む談話構造木を構築すればよいということと同義である。

5 談話構造木の構築

5.1 談話構造木

本章では談話構造木の構築アルゴリズムについて述べる。談話構造木は文をノードとし、談話中の話題の推移を表す木構造である。談話構造木中では、近い話題を持つ、つまり結束性の高い文同士が隣接する。また、ルートノード（談話構造木の根に当たるノード）は談話の先頭文であり、各ノード文は自身よりも前に出現した文のうちのただ一つを親ノード文として持つ。

談話構造木構築アルゴリズムでは、先頭から順に処理対象文とし、先頭文はルートノードに、以降の文は自身よりも前に出現した文すべてを接続候補文とし、最も結束性の高い接続候補文に接続する。文同士の結束性を測るための指標として、センタリング理論（第 4 章）における TRANSITION の種類により結束性の度合いを表すセンタリング得点、特定の表層パターンの組み合わせに着目した表層パターン得点、文間距離を用いた文間距離得点の 3 つの指標の和を用いる。以下でそれぞれについて説明する。

5.2 センタリング得点 C_p

センタリング理論では、文間の焦点の推移に着目した TRANSITION を求めることにより文間の結束性の高さを評価できる。接続候補ノード文との TRANSITION を求め、TRANSITION の種類により結束性が高いものほど高得点とする。ここで TRANSITION の判定には、対象知識も含める。つまり、接続候補文の C_b と処理対象文の C_b の関係が対象知識上の関係に該当すれば TRANSITION は CONTINUATION か RETAINING になる。ただし、 C_b が同一概念や同義概念ではなく対象知識から決定される概念を持つ語の場合は、同一概念や同義概念の場合よりも得点を低くするように、 C_b の決定タイプにより重み付けする。センタリング得点 C_p は以下の式に従う。

$$C_p = T_p \times w_{Cb} \quad (2)$$

T_p : TRANSITION 得点

w_{Cb} : C_b 重み

T_p は TRANSITION に応じた得点であり、結束性の高いものほど高得点になるように設定する。 w_{Cb} は C_b の決定タイプに応じた重みであり同一概念や同義概念を 1.0 とし、優先度に応じた値を設定する。表 7 に TRANSITION 得点 T_p 、表 8 に C_b 決定タイプによる重み w_{Cb} の例を示す。

5.3 表層パターン得点 E_p

接続候補文と処理対象文が特定の表層パターンの組み合わせ（例えば、接続候補文の表記が「最初に」で始まり、処理対象文の表記が「つぎに」で始まる）を持てば、関連の高い話題を

表 7 TRANSITION 得点 T_p の例

TRANSITION	T_p
CONTINUATION	90
RETAINING	60
SHIFTING	30
NOTHING	0

表 8 C_b 決定タイプによる重み w_{Cb} の例

C_b タイプ	w_{Cb}
EQUAL	1.0
ATTRIBUTE	0.6
LOWER	0.6
SIMILAR	0.3

表 9 表層パターン組み合わせの例

接続候補文の表層パターン	処理対象文の表層パターン
まず、はじめに、最初に	次に、さらに、続いて、それから
次に、さらに、続いて、それから	次に、さらに、続いて、それから
次に、さらに、続いて、それから	最後に
第 1 に、1 つ目に、…	第 2 に、2 つ目に、…

順序立てて提示していると考えられ、結束性は高いと判断し、表層パターン得点 E_p を付与する。表層パターンが現れる場合は接続先として確定的な場合であり、 E_p はセンタリング得点の CONTINUATION の場合などよりも高い値に設定する。このような表層パターンの組み合わせの例を表 9 に示す。

5.4 文間距離得点 D_p

発話者は、読者にわかりやすいように、話題を徐々に変化させながら談話を構築する。そのため近接する文は関連性の高い話題を持つことが自然である。そこで近接する文ノードほど結束性が高いと判断し、文間距離に反比例する得点を与える。

$$D_p = D_{p_{\max}} \left(1 - \frac{d}{i} \right) \quad (3)$$

d : 文間距離

i : 処理対象文番号

$D_{p_{\max}}$: 距離得点上限值

ここで、 d は接続候補文と処理対象文との文間距離、 i は処理対象文番号、 $D_{p_{\max}}$ は文間距離得点上限值である。 $D_{p_{\max}}$ は TRANSITION 間の得点を大きく超えない値が望ましい。つまり表 7 の場合 CONTINUATION が 90、RETAINING が 60 であるので $D_{p_{\max}}$ は 30 とした。

5.5 談話構造木構築アルゴリズム

以上の 3 つの指標の合計得点を結束性得点とし、接続候補文それぞれに対して求め、最高得点を持つ文を親ノードとしてリンクを張る。談話構造木構築アルゴリズムを図 3 に示す。

以下の文章を例に説明する。

N 自動社は 19 日、新車発表会を開催。N 自動車としては初となる電気自動車を公開した。電気自動車は、走行時に CO_2 を排出しないとして注目を集めている。販売を伸ばしているハイブリット車への対抗として、N 自動車は巻き返しを図る考えだ。

```

for(すべての文を処理対象文Siに)
{
    if(先頭文であれば)
    {
        ルートノードにする
    }
    else
    {
        for(処理済のすべての文を接続候補文Sjに)
        {
            結束性得点P[Sj] += 表記パターン得点Ep
            結束性得点P[Sj] += センタリング得点Cp
            結束性得点P[Sj] += 文間距離得点Dp
        }
        最高の結束性得点P[Sj]を持つ接続候補文Sjとリンクを張る
    }
}

```

図 3 談話構造木構築アルゴリズム

図 4 談話構造木の構築例

ここで TRANSITION 得点 T_p は表 7, 文間距離得点上限值 $D_{p_{max}}$ は 30 とし説明する。まず 1 文目をルートノードにする。2 文目では、接続候補文は 1 文目のみであるので 1 文目とリンクを張る。3 文目では、1 文目と 2 文目が接続候補文になるが、それぞれに対し TRANSITION を求めると、1 文目に対しては NOTHING, 2 文目に対しては SHIFTING となる。表層パターンはなく、距離得点を加算する。結束性得点は 1 文目 15, 2 文目 52.5 となり、2 文目が最も結束性得点が高くなるので、2 文目とリンクを張る。4 文目は、1 文目から 3 文目までが接続候補文となる。

それぞれに対して、TRANSITION を求める。図 4 は各接続候補文に対して TRANSITION を求めた状態である。1 文目と 2 文目に対しては CONTINUATION, 3 文目に対しては SHIFTING となる。表層パターンはなく、距離得点を加算する。結束性得点は、1 文目 97.5, 2 文目 105, 3 文目 52.5 となり最も得点の高い 2 文目とリンクを張る。

6 文間接続関係の判定

6.1 文間接続関係

構築された談話構造中のすべてのリンクに文間接続関係を付与する。文間接続関係はリンクで直接接続された 2 文間の接続関係を表す。何種類の間接関係を定義するかは、研究者により異なっているが (6 種類～11 種類)、本研究では、黒橋ら (黒橋, 長尾 1994) の 11 種類の結束関係 (並列, 対比, 主題連鎖, 焦点—主題連鎖, 詳細化, 理由, 原因—結果, 変化, 例提示, 例説明, 質問—応答) を参考に、事例 13 文章 112 文に試行する過程で、意味解析結果の語意、深層格、モダリティや、対象知識や主題や話題を基にした計算可能性と要約や質問応答などの応用での必要性から、表 10 の 9 種類を定義し、利用することにした。

文間接続関係は基本的には、親ノード文から子ノード文への関係である。つまり表 10 での S_i が親ノード文であり、 S_j が子ノード文である。例えば、文間接続関係が「詳細化」といった場合は親ノード文での話題に関して子ノード文でさらに詳細な情報が記述されているということである。ただし、「原因結果」と「逆接」関係に関しては逆向き関係が存在する。

なお、黒橋らとの比較で言えば、判定の明確性を維持し 6.2 節の文間接続関係判定ルールを容易に構築できるようにするために、対比を逆接として、変化を遷移として再定義した。また同様の目的から主題連鎖と焦点—主題連鎖の違いを、展開と転換に分離再定義した。一方、理由と例説明は出現頻度が低いので今回の分類では削除した。具体的な事例では、これらは詳細化に分類されることが多いと思われる。

表 10 文間接続関係一覧

詳細化	S_i で述べられた話題を継続し、その詳しい内容が S_j で述べられている
展開	S_j で S_i に関連する新たな話題を提示する
原因—結果	S_i が原因でその結果として S_j が起る
逆接	S_j が S_i から予測される結果に反する事象である
転換	S_j で S_i に関連のない話題が提示されている
遷移	時間変化に伴い、 S_i の事象の後に S_j で記述された事象が起る
並列	S_i と S_j が並列関係にある
例提示	S_i で述べられた事象や状態の具体例の項目が S_j で提示される
質問—応答	S_i の問題に対して S_j で答えが示される

6.2 文間接続関係判定ルール

文間接続関係の判定には、文間接続関係判定ルールを用いる。文間接続関係判定ルールは以下の形式を持つ。

- 接続関係名
 - － 9 種類の接続関係（表 10）のうち 1 つ
- 条件部
 - － 接続詞
 - － 主題
 - － モダリティ・テンス・アスペクト
 - － 対象知識
 - － 構文・意味情報
- 得点
 - － 接続関係名への確信度

文間接続関係判定ルールは接続関係名と条件部、得点から構成される。条件部は接続詞などの表層表現や主題の推移、文の認識や話者の態度を表すモダリティ（推量、疑問など）、時制を表すテンス・アスペクト、対象知識、構文・意味情報を用いた論理式である。ここで主題とは、その文が何について述べられているのかを示すもので、一般的には、助詞「は」で示される文節のことを指す。本研究では、助詞「は」で示される文節のほかに助詞「も」や読点「、」で示される主語格も対象とする。なおモダリティテンスアスペクト、接続詞、構文・意味情報は、モダリティ解析機能が加わった意味解析システム Sage (梅澤, 西尾, 松田, 原田 2008; 梅澤, 加藤, 松田, 原田 2009) で解析された情報に基づく。ルールの例とその適合例を図 5～図 7 に示す。

図 5 文間接続関係判定ルールの例 1

接続関係名: 逆接
条件部: 子ノード文の先頭に接続詞「しかし」
得点: 100

図 6 文間接続関係判定ルール例 2

接続関係名: 詳細化
条件部: 子ノード文の主題が親ノード文の主題の部分/属性概念
得点: 50

図 7 文間接続関係判定ルール例 3

図 5 の例では、親ノード文の主題以外の文節「電気自動車」が子ノード文の主題として現れている。親ノード文の話題「N 自動車」から新たな話題「電気自動車」に展開されている。図 6 では、子ノード文が接続詞「しかし」で始まっている。接続詞「しかし」は逆接を表す一般的な接続詞であり、このような接続詞が現れる場合は、接続関係が明確に現れている場合であり、高得点を与える。図 7 では親ノード文の主題「電気自動車」が子ノード文で部分／属性である「充電時間」になっている。この場合では親ノード文の話題を引き続きつつより詳細な内容の説明へ移行していると考えている。

ルールの一覧を表 11 に、ルールで使われる接続詞の一覧を表 12 に、それを用いる文間接続関係名ごとに示す。ここで (R) と書かれているものは逆向き関係を表している。

表 11 文間接続関係判定ルール一覧

接続関係	条件部	得点
詳細化	子ノード文の主題が親ノード文の主題と同義	50
詳細化	子ノード文の主題が親ノード文の主題の部分／属性	50
詳細化	子ノード文の主題が親ノード文の主題の下位	50
詳細化	子ノード文が詳細化の接続詞で始まる	50
詳細化	子ノード文の主語が省略	10
詳細化	子ノード文の文末が形容詞節または形容動詞節, 名詞節	10
詳細化	デフォルト得点	10
展開	子ノード文の主題が親ノード文の主題以外の要素と同義	40
展開	子ノード文の主題が親ノード文の主題以外の要素の部分／属性	40
展開	子ノード文の主題が親ノード文の主題以外の要素の下位	40
展開	子ノード文が展開の接続詞で始まる	50
展開	子ノード文の主題が指示詞	50
原因—結果	子ノード文が原因—結果の接続詞で始まる	100
原因—結果	子ノード文の文末のモダリティが推量・勧誘・依頼・意志・希望	50
原因—結果	子ノード文の文末が「思う・考える」	50
原因—結果 (R)	子ノード文が原因—結果 (R) の接続詞で始まる	100
原因—結果 (R)	子ノード文の文末が「から・ため・による」	80
逆接	子ノード文が逆接の接続詞で始まる	100
逆接	子ノード文が理由を尋ねる文	80
逆接	子ノード文と親ノード文の主題が異なり文末節の主辞が同義で片方が否定形	80
逆接 (R)	子ノード文の文末が「のに」	80
転換	子ノード文が転換の接続詞で始まる	100
転換	子ノード文の主題が親ノード文に含まれない	30
遷移	子ノード文が遷移の接続詞で始まる	50
遷移	子ノード文と親ノード文のテンスが過去で子ノード文のアスペクトが完結相	50
遷移	子ノード文の親ノード文の文末が動詞節	20
遷移	子ノード文と親ノード文に時間に関する記述がある	80
遷移	子ノード文とのみに時間に関する記述がある	20
遷移	子ノード文に主題がなくガ格がある	10
並列	子ノード文が並列の接続詞で始まる	100
並列	子ノード文と親ノード文が基本形以外で同一のモダリティを持つ	80
例提示	子ノード文が例提示の接続詞で始まる	100
例提示	子ノード文の文末が断定節である	20
例提示	子ノード文の主題が引用である	50
質問—応答	子ノード文が質問—応答の接続詞で始まる	100
質問—応答	親ノード文のモダリティが疑問で子ノード文のモダリティが疑問ではない	60

表 12 接続詞一覧

接続関係	接続詞
詳細化	そして、それで、それと、あと、それに、そのうえ、そのうえに、おまけに、そればかりか、ひいては、しかも、それも、すなわち、つまり、いいかえると、いわば、いってみれば、代わりに、その代わり、その代わりに、いな、そうではなく、そうではなくて、そうじゃなくて、というより、というよりも、というか、むしろ、このように、以上、要するに、こうして、結局、このようにして、ただし、ただ
展開	このように、以上、要するに、こうして、結局、このようにして、もっとも、ちなみに、なお、いずれにしても、どっちみち、どっちにしても、何にしても、何はともあれ、とにかく、ただし、ただ
原因—結果	だから、それで、そのため、したがって、従って、ゆえに、故に、それ故、それ故に、よって、そこで、そのせいか、なので、で、それなら、なら、それでは、では、そうしたら、すると、そうすると、とすると、してみると、そうなると、こうなると、となると、そうすれば、そうして、こうして、さもないと、さもなければ、そうでなければ、そうでないと、そうしないと、そうしなければ
原因—結果 (R)	何故なら、なぜなら、何故ならば、なぜならば、何故かという、なぜかという、どうしてかという、というのは、というの、だって、それというの、だって、ついては、何となれば、その結果、その理由は
逆接	しかし、だけど、でも、ところが、それが、それなのに、そのわりに、そのくせ、といっても、とはいっても、とはいえ、そうはいっても、それに対して、反対に、反面、逆に、一方、他方、かといって、が、けど、けども、けれど、けれども、さりとて、されど、しかしながら、それでいて、それにしても、それにもかかわらず、だからと言って、だが、だけれども、ですが、ですけれども、ですけれど、ですけれども、とはいっても、なのに、にもかかわらず、反面、しかるに、一方、それにひきかえ、それに比べ、それに対して、これに対して
転換	ところで、さて、それにしても、話は飛ぶけど、話は変わるけど、それはそうと、それより、ときに、そういえば、それとは別に
遷移	そして、それで、それと、あと、それに、そのうえ、おまけに、そればかりか、ひいては、しかも、それも、代わりに、その代わり、まず、はじめに、最初に、次に、続いて、ついで、次いで、それから、また、さらに、他に、最後に、このように、こうして、結局、このようにして、やがて、かくして、かくて、そしたら、最終的には
並列	第 1 に、第 2 に、1 つめに、2 つめに、まず、はじめに、最初に、次に、続いて、ついで、それから、また、さらに、他に、最後に
例提示	実際、事実、たとえば、例えば、特に、とりわけ、なかでも
質問—応答	その答は

6.3 文間接続関係判定アルゴリズム

つぎに文間接続関係判定のアルゴリズム図 8 について述べる。文間接続関係判定のアルゴリズムでは、談話構造木に存在するすべてのリンクに対し、1 つずつ順に文間接続関係を付与する。まず、リンク 1 つに対し接続関係それぞれに対応する 9 つの適合得点を用意する。そして

```

for(すべての文間リンクを対象に)
{
    for(すべてのルールを対象に)
    {
        if(ルールの条件部に適合)
            対応する接続関係の適合得点 += ルールの得点
    }
    リンクの接続関係 = 最高の適合得点を持つ接続関係
}

```

図 8 文間接続関係判定アルゴリズム

表 11 のルール 1 つずつ、条件部に適合するか判定する。適合した場合には、ルールの得点を対応する接続関係の適合得点に加算する。そしてすべてのルールについて適合判定を行った後、最高の適合得点を持つ接続関係を文間接続関係として決定する。

7 評価実験と考察

7.1 評価実験

本研究で作成した DIA で談話構造解析を行った様子を図 9 に示す。左上のテキストボックスに解析するテキストを入力。実行すると左下に談話解析の結果である談話構造木が表示される。各入力文はノードで表示され、接続関係にある文同士が親ノードから子ノードへのアークで接続されている。また文間接続関係を表すラベルがアーク上に付与される。例えば、4 文目「日航株は 1 円の値上がりでも、大幅な上昇率となる。」と 6 文目「ただ、売り抜けることができないと、全額損失になる。」では「株」と「損失」の部分／属性関係を基に、4 文目と 10 文目「市場では「一か八かのギャンブル相場入りした」(大手証券関係者)との声が出ている。」では「株」と「相場」の部分／属性関係を基に正しく接続先の決定ができています。また談話構造木の左下の部分木は、詳細化として株価が 1 円近辺になった時の具体的な値動きについての話題で構成されている。一方、右下の部分木では日航株が話題となっている。この結果を、自動要約に用いるには、根の文と本文章の最後に近い葉の文に至る道上から要約文を作ることが有効のように思われる。

本研究はこれまでに述べてきたように、談話構造木中の接続先判定においても、文間接続関係の決定においても、筆者らの事例調査と経験によって作成したルールによっている。このルールを作成するにあたって筆者らが本論文で述べて談話構造木の作成の基本アルゴリズムをベースに手作業で事例の文章の談話構造木を作成する過程において、アルゴリズム中の表 8 に示した Cb 決定タイプによる重み w_{Cb} などのいくつかの数値パラメータを決定していった。このシス

テム作成時に事例として利用したのは Web 上のニュース記事から抽出した 13 文章 112 文である。これらを対象にクローズドテストを行った。

各種パラメータや得点は出来るだけ正しい談話構造が求まるよう人手で調整した。実験に用いたパラメータは、TRANSITION 得点の CONTINUATION は 90, RETAINING は 55, SHIFTING は 30, NOTHING は 0, 距離得点上限値は 33, Cb 決定タイプによる重みは、部分／属性（共起辞書から抽出）は 0.6, 部分／属性（上位概念ペア）は 0.3, 上位一下位 0.6, 類義 0.3 とした。文間接族関係の判定ルールの得点は表 11 のとおりである。実験結果を表 13 に示す（分母が 99 なのは接続関係の数が $112 - 13 = 99$ だからである）。

接続先文の正解率は談話構造木の構築で正しい親ノード文を選んだ割合を表している。文間

図 9 DIA の実行例

表 13 実験結果

接続先文の正解率	文間接続関係の正解率	全体の正解率
77.8% [77/99]	85.7% [66/77]	66.7% [66/99]

表 14 実験結果

接続先文の正解率	文間接続関係の正解率	全体の正解率
82.1% [96/117]	81.3% [78/96]	66.7% [78/117]

接族関係の正解率は談話構造木の構築で正しい接続先が選ばれたものの中での文間接族関係の正解率である。全体の正解率は、正しい接続先を選びかつ正しい文間接族関係を選んだ割合である。

さらに、Web から得た上記とは異なる 12 文章 129 文を対象にオープンテストを実施した。正解の判定は筆者らが所属する研究室の自動要約を研究している筆者らとは別の学生に依頼した。その結果を表 14 に示す。

7.2 考察

実験の結果、談話構造木の構築では、同一概念や同義概念が含まれる場合は正しく接続先が特定されたが、部分／属性概念、上位／下位概念など対象知識から得られる概念関係を用いた接続先の決定は表 8 に示した Cb 決定タイプによる重み w_{Cb} に依存し、誤る事例はこの数値が原因であることが多かった。現在 w_{Cb} は経験的に定めているが、今後より多くの正解事例を手作業で作成し機械学習によって定めることによってより精度の高い接続先の決定が行えると思われる。

一方、部分／属性の判定精度を向上できると接続先の決定精度も向上する。部分／属性の判定は、共起辞書を用いた規則の場合は高い確率で正しい関係を導いている。上位概念ペアを用いた規則の場合は、部分／属性の上位概念として選定した概念（表 3 の右列に列挙した概念）がシソーラスでの位置で根から近（高層）いことが原因で、少し精度が低くなっていることがある。より下層の概念（その分概念数は多くなるが）を上位概念として選定すれば誤りを除くことができ、精度は向上する。また 2 語間だけでみた場合は正しい部分／属性関係であっても、文脈上ではそうではない場合もあった（例：属性が「関係者」でそれを持つ対象が「法人」や「施設」など文脈上に複数の候補がある場合）。

文間接続関係の判定は、高い精度を実現したが、詳細化と遷移間の分離精度が少し低かった。さらにセンタリング理論で同義概念など概念の関係をみるとき文節を単位に比較したが、文節の区切りに表現上のゆらぎがあり、さらに大きい単位でみないと正しく比較できない場合があった（例：「通信技術」と「通信の技術」。後者は 2 文節）。

本研究の応用については、現在、自動要約、照応解析、質問応答に用いることを試行している。自動要約においては、要約の種となる重要語を選定する際の得点として、談話構造木において原文の最初の文から最後の文に至る路上の文に含まれる場合に加点する手法で、話の主要

な流れを漏れなくカバーする要約文の生成を期待できる。ゼロ代名詞の先行詞の判定を行う照応解析では、先行詞候補を探索する範囲を談話構造木において先頭の文からゼロ代名詞のある文に至る経路上の文に高い精度で絞り込むことができると期待している。質問応答システムでは、文間接続関係を用いて、理由や原因の回答抽出が容易になることや、質問文と類似度の高い照応文という知識文（新聞や web 中の回答を得ようとする知識ソース中の文）から実際に回答を含む回答文を探索する範囲を談話構造木中で照応文を含む経路上に限定できることなどが期待できる。

8 おわりに

本研究では、センタリング理論と対象知識に基づき、談話における話題の推移を、正確に捉える談話構造解析の手法を示した。対象知識を用いることで、同一概念や同義概念だけでなく部分／属性や上位／下位概念への話題の展開を考慮し、表記上の手掛かりがない文章でも焦点の推移を的確にとらえることが可能となった。また、センタリング理論により、焦点の推移と連続性を評価したことで、結束性の高い談話構造木を構築することができた。文間接続関係判定ルールと対象知識判定規則の精緻化や、機械学習などによる各種パラメータの高精度な決定、辞書の整備、表記のゆれに対する解決などを行えばさらに高い精度を実現できると思われる。

謝 辞

本研究を進めるにあたって有意義なコメントを頂いた青山学院大学原田研究室の皆様へ感謝いたします。特に、久保田裕章氏には接続先の決定方法や文間接続関係の選出の議論に参加して頂き有意義な意見を頂いた、また西尾公秀氏には丁寧なオープンテストを実施して頂いた、深く感謝いたします。

参考文献

- Grosz Barbara J, Weinstein Scott and Joshi Aravind K (1995). "Centering: A Framework for Modeling the Local Coherence of Discourse." *Association for Computational Linguistics*, 21, pp. 203–225.
- 原田実, 尾見孝一郎, 岩田隆志, 水野高宏 (1999). 日本語文章からの意味フレーム自動生成システム SAGE (Semantic frame Automatic GEnerator) の開発研究. 人工知能学会第 13 回全国大会論文集, pp. 213–216.

- 原田実, 水野高宏 (2001). EDR を用いた日本語意味解析システム SAGE. 人工知能学会論文誌, **16** (1), pp. 85–93.
- 原田実, 田淵和幸, 大野博之 (2002). 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価. 情報処理学会論文誌, **43** (9), pp. 2894–2902.
- (株) 日本語電子辞書研究所 (1995). EDR 電子化辞書仕様説明書 (第 2 版).
- 加藤直人, 森元逞 (1995). 統計的手法による談話構造解析. 情報処理学会第 51 回全国大会, pp. 99–100.
- 加藤裕平, 古川勇人, 蒲生健輝, 韓東力, 原田実 (2005). WEB 検索による知識文の獲得と意味グラフ照合推論による質問応答システム Metis. 情報処理学会第 67 回全国大会論文集, 1G-06, 第 2 分冊, pp. 11–12.
- 黒橋禎夫, 長尾真 (1994). 表層表現中の情報に基づく文章構造の自動抽出. 自然言語処理, **1** (1), pp. 3–20.
- 黒橋禎夫, 長尾真 (1998a). 日本語形態素解析システム JUMAN version 3.61. 京都大学大学院情報学研究科.
- 黒橋禎夫, 長尾真 (1998b). 日本語構文解析システム KNP 使用説明書 version 2.0b6. 京都大学大学院情報学研究科.
- Marcu Daniel and Echiabi Abdessamad (2002). “An Unsupervised Approach to Recognizing Discourse Relations.” *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 368–375.
- 南旭瑞, 原田実 (2002). 語意の類似性を用いた照応解析システムの開発 Anasys. 情報処理学会第 64 回全国大会論文集, 3M-06 第 2 分冊, pp. 53–54.
- 横山憲司, 難波英嗣, 奥村学 (2003). Support Vector Machine を用いた談話構造解析. 情報処理学会研究報告 自然言語処理研究会報告, 23, pp. 193–200.
- 柴田和秀, 黒橋禎夫 (2005). 隠れマルコフモデルによるトピックの遷移を捉えた談話構造解析. 言語処理学会第 11 回年次大会, pp. 109–112.
- Sporleder Caroline and Lascarides Alex (2005). “Exploiting Linguistic Cues to Classify Rhetorical Relations.” *Proceedings of Recent Advances in Natural Language Processing*, pp. 532–539.
- 田中信彰, 面来道彦, 野口貴, 矢後友和, 韓東力, 原田実 (2006). 意味解析を踏まえた自動要約システム ABISYS. 言語処理学会論文誌, **13** (1), pp. 143–164.
- 梅澤俊之, 西尾華織, 松田源立, 原田実 (2008). 意味解析システム SAGE の精度向上とモダリティの付与と辞書更新支援系の開発. 言語処理学会第 14 回年次大会発表論文集, E3-1, pp. 548–551.
- 梅澤俊之, 加藤大知, 松田源立, 原田実 (2009). 意味解析システム SAGE の精度向上—モダリティと副詞節について—. 情報処理学会第 191 回自然言語処理研究会, pp. 1–8.

山本和英, 斉藤真実 (2008). 用例利用型による文間接続関係の同定. 自然言語処理, **15** (3), pp. 21-51.

略歴

梅澤 俊之：2008 年青山学院大学理工学部情報テクノロジー学科卒業. 2010 年青山学院大学大学院理工学研究科理工学専攻知能情報コース博士前期課程修了.

原田 実：1975 年東京大学理学部物理学学科卒業. 1980 年東京大学理学系大学院博士課程修了. 理学博士. (財)電力中央研究所研究員を経て, 1989 年青山学院大学理工学部経営工学科助教授に就任, 2000 年より同情報テクノロジー学科教授. 1986 年電力中央研究所経済研究所所長賞. 1992 年人工知能学会全国大会優秀論文賞. 2008 年青山学院学術褒賞. 主たる研究は, 自動プログラミング, 意味理解, 自動要約, 質問応答, テキストマイニング, 対話応答などにおいて実利用可能な技術開発に従事. 情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE, ACM 各会員.

(2010 年 4 月 20 日 受付)

(2010 年 9 月 3 日 再受付)

(2010 年 10 月 4 日 採録)