

2019年度 博士学位論文

日本語談話関係認識のための
理論とコーパス構築

お茶の水女子大学大学院
人間文化創成科学研究科 理学専攻

金子 貴美

2020年3月

Abstract

One of the fundamental goals of natural language processing is to develop a computational system to understand the meaning of texts within contexts. Such meanings involve information related to the structure between sentences, which is not derivable by sentential semantics alone; it is necessary to analyze the semantic relations between sentences and clauses. Such semantic relations have been called “discourse relations”, and studied in the research field known as discourse theory in both natural language processing and theoretical linguistics.

Since discourse relations and temporal relations interact with each other, we should take them into consideration of the definition of discourse relations. However, dependencies between discourse relations and temporal relations are not fully analyzed in previous researches on discourse relations.

Based on the above observations, the aim of this study is, firstly, to sort out the dependencies of the intentions of speakers/writers on the discourse relations and secondly to establish a method for uniquely determining the discourse relation for each specific sentence/clause pair. In order to achieve this aim, I establish the decision procedure for judging discourse relations reflecting the dependencies between discourse relations and temporal relations. More concretely, I first analyze the dependencies between eventuality, modality, tense and aspect, temporal relation, and discourse relation, based on which I define the annotation schema and the decision procedure for discourse relations. I performed an annotation experiment following my annotation schema and the decision procedure. By calculating the inter-annotator agreement and conducting error analysis, I verify whether I could sort out the relations between the intentions of speakers/writers and determine the unique judgment. Finally, I discuss which part of the procedure is left unclear as future work.

概要

自然言語処理研究の究極的な目標の1つは「計算機に文章・文脈を理解させること」である。文章は、それを構成する各文を意味解析しただけでは導出できない、文間の構造に関わる意味を持っているため、上記目標を実現するには、文間や節間にある意味的な繋がり関係を解析する必要がある。このような、文や節といった談話片間の意味的な関係のことを談話関係と呼ぶ。

談話関係と時間関係は相互に依存しているため、そのような依存関係を考慮して談話関係の定義がなされるべきである。しかしながら、談話関係に関するいずれの関連研究も、談話関係と時間関係の依存関係を反映した設計となっているとは言い難い。そのような設計ができていない一因として、いずれの関連研究も解釈が揺れる「話し手や書き手の意図に関する関係」と、解釈が一意に定まる関係とを十分に選り分けられていないことがある、と推察される。

上記の問題点を踏まえ、本研究では、話し手や書き手の意図に関する関係と、解釈が一意に定まる関係とを選り分け、後者について、時間関係との依存関係を反映した、談話関係の決定手順を確立することを目指し、談話関係を再定義する。具体的には、上記の実現のためにまず、出来事 (eventuality)、モダリティ (modality)、時制 (tense)、時間関係 (temporal relation) と、談話関係との依存関係の整理を行い、それを元に談話関係の分類体系と客観的な談話関係の決定手順を定義する。その後、定義した分類体系と決定手順に従ってアノテーション実験を行い、一致率の算出やエラー分析を行うことにより、「解釈が一意に定まる関係のみに絞って談話関係を定義できたか」を検証し、談話関係の決定プロセスのどの部分を明らかにできたか、および、どの部分に課題が残るかを議論する。

目次

第 1 章	序論	1
第 2 章	関連研究	5
2.1	修辞構造理論 (RST)	5
2.2	Discourse GraphBank (DGB)	8
2.3	Penn Discourse Treebank (PDTB)	10
2.4	分節談話表示理論 (SDRT)	15
2.5	日本語における談話関係アノテーションの研究	19
2.6	関連研究における課題と本研究の位置付け	21
第 3 章	談話関係アノテーション手法の提案	24
3.1	作業概要	24
3.2	節の述部のアノテーション	26
3.3	出来事・モダリティのアノテーション	33
3.4	述部間の接続先の判定方法	39
3.5	談話関係アノテーション	42
3.6	提案手法の有効性に関する定性的評価	63
第 4 章	談話関係アノテーションの実験結果と分析	68
4.1	実験設定	68
4.2	談話関係アノテーションの結果	70
4.3	エラー分析	73
第 5 章	結論	76
5.1	まとめ	76
5.2	今後の課題・展望	77

謝辭	78
参考文献	80
発表文献一覽	83

表目次

2.1	RST-DT[5]における談話関係の定義の例	6
2.2	RST-DT[5]における関係一覧 (談話関係と Schemata の関係)	7
2.3	DGB[6]における関係一覧	9
2.4	PDTB-3.0[10] の関係ラベルの階層関係	11
2.5	PDTB-3.0[10] の関係ラベル一覧	12
2.6	PDTB-3.0[10] の関係ラベル一覧 (続き)	13
2.7	PDTB-3.0[10]における談話関係の定義の例	14
2.8	SDRT[4]における談話関係一覧 (Coordinating Relations)	18
2.9	SDRT[4]における談話関係一覧 (Subordinating Relations)	19
2.10	河原らの談話関係一覧 [13, 14, 15]	21
3.1	eventuality/modality 分類表	36
3.2	視点時と出来事時と Temporal Relations の対応	49
3.3	(13)における、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果	64
3.4	(14)における、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果	65
4.1	本研究における関係ラベルの分布	71
4.2	関係ラベル名と省略表記の対応表	71
4.3	本研究における関係ラベルと出来事情報の共起頻度分布	72
4.4	本研究における関係ラベルと時制情報の共起頻度分布	72

目次

1.1	例 (5) の出来事の時区間の配置	2
1.2	例 (6) の出来事の時区間の配置	2
2.1	RST の談話構造の例 (Carlson et al.[5] p.25 Figure 3 より)	5
2.2	DGB の談話構造の例 (Wolf et al.[6] p.5 Figure 1 より)	8
2.3	(π_1)-(π_5) の意味表示 (Asher et al.[4] p.147 Figure 4.5 より)	17
2.4	節ペアに対する談話関係の有無の判定のタスク画面 (河原ら [15] p.820 図 1 より)	20
2.5	談話関係があると判定された節ペアに対する談話関係タイプの判定のタスク画面 (河原ら [15] p.820 図 2 より)	20
3.1	24
3.2	述部に活用形・時制ラベルと出来事の情報付与した例	25
3.3	接続先判定と談話関係ラベルの付与を行った例	25
3.4	述部判定の状態遷移図	27
3.5	例 (40) の出来事の時区間の配置	33
3.6	例 (41) の出来事の時区間の配置	33
3.7	例 (42) の出来事の時区間の配置	34
3.8	述部「濡れている」の eventuality の変化	35
3.9	(a) の例	40
3.10	(a') の「関係節内部の述部同士が係り受け関係にある場合」の例	40
3.11	(a') の「関係節内の述部と関係節外の述部」の例	40
3.12	(b) の例	40
3.13	(c) の例	41
3.14	一方の文にのみモダリティがある場合の例	41
3.15	(モダリティがある文同士で接続する例	41

3.16	単文の時間配置	43
3.17	例 (52) の時間配置 1	43
3.18	例 (52) の時間配置 2	44
3.19	2 つの談話関係ラベルが付与される例	45
3.20	(53) の時間配置の例 1	46
3.21	(53) の時間配置の例 2	46
3.22	浅原ら [28] の時間関係ラベルと出来事の時間配置の対応	48
3.23	(41) の時間配置の例	50
3.24	(54) の時間配置	51
3.25	(55) の時間配置	51
3.26	(40) の時間配置	52
3.27	(56) の時間配置	53
3.28	(57) の時間配置	54
3.29	(58) の時間配置	54
3.30	(59) の時間配置	55
3.31	(62) の時間配置	57
3.32	(63) の時間配置	57
4.1	brat における述部アノテーションの様子	69
4.2	brat 上で述部に付与された eventuality/modality の情報の例	69
4.3	brat における談話関係アノテーションの例	70
4.4	「どの init-eventuality、最終的な eventuality/modality、時制が対になっているかの情報」の例	70

第 1 章

序論

自然言語処理研究の究極的な目標の 1 つは「計算機に文章・文脈を理解させること」である。近年も、文章読解 (reading comprehension) や質問応答 (question answering) などといった文章・文脈の理解を伴うタスクにおける研究開発が、学术界、産業界問わず盛んに行われている。文章は、それを構成する各文を意味解析しただけでは導出できない、文間の構造（ひいては文章全体の意味）に関わる意味を持っているため、上記目標を実現するには、文間や節間にある意味的な繋がりを解析する必要がある。このような、文や節といった談話片間の意味的関係のことを**談話関係**と呼ぶ。例えば、以下の 2 つの文章を考える。

- (1) a. 花子は落とし穴に 落ち、
b. 太郎は花子を 助けた。
- (2) a. 太郎が花子を 押した から、
b. 花子は落とし穴に 落ちた。

(1)(2) における従属節と主節の間関係は、(3)(4) のように文間関係としても現れる。このように、2 文に分けられる（書き換えられる）節間、ないしは文間関係のことを談話関係という。

- (3) a. 花子は落とし穴に 落ちた。
b. 太郎は花子を 助けた。
- (4) a. 花子は落とし穴に 落ちた。
b. 太郎が花子を 押した からだ。

ここで、(3)(4) に示す例における 2 文間の時間関係は異なっている。(3) では、「落ちた」という出来事と「助けた」という出来事が起きた順序は文の順序と合っている一方、(4) では、「押した」こ

とにより「落ちた」と解釈できるため、出来事の成立順序と文の順序は合っていない。単に文の順序に沿って、文の意味解析結果を繋ぎ合わせて文章の解析結果とした場合、(3)(4)の時間配置を区別することができない。しかしながら、談話関係が解析できれば、これらを区別することができる。このように談話関係の理解は、文章内の出来事間の時間的・空間的 (spatio-temporal) な関係を捉える上で重要な役割を果たすため、自然言語理解に必須のモジュールであると考えられる。

一方、談話関係は (3)(4) のように談話関係が決まると時間関係が特定されるものだけでなく、時間関係が決まると談話関係が特定されるものがある。次の2つの文章について考える。

- (5) a. カーテンを 開けた。
b. 一面の銀世界が 広がっていた。
- (6) a. カーテンを 開けた。
b. 一面の銀世界を 見下ろした。

(5)(6) に示す例における2文間の時間関係も異なっている。(5) では、図 1.1 のように「開ける」動作の成立時に「広がっている」状態が成立していたと解釈できる。



図 1.1 例 (5) の出来事の時区間の配置

一方 (6) では、図 1.2 のように「開ける」の動作の後、「見下ろす」動作を行ったと読み取れる。



図 1.2 例 (6) の出来事の時区間の配置

このように、談話関係と時間関係は相互に依存しているため、そのような依存関係を考慮して談話関係の定義がなされるべきであると考えられる。

談話関係の理解の必要性は古くから認識されており、談話関係の理論やコーパス構築に関する先駆的な研究がいくつか存在する [1, 2, 3, 4]。しかしながら、談話関係に関するいずれの関連研究も、談話関係と時間関係の依存関係を反映した設計となっているとは言い難い。まず、修辞構造理論 (Rhetorical Structure Theory ; RST)[1] と Discourse GraphBank(DGB)[2] における談話関係は、客観的な付与基準が定義されておらず、読み手の判断に委ねられている。それ故に、再現性が担保されない、アノテータの判断のプロセスを結果から遡って分析することが難しい、などといった問題がある。一方で、Penn Discourse Treebank(PDTB)[3] は、接続表現に基づいた RST や DGB より客観的な判断基準を定義しているが、談話関係とその判断基準の定義が接続表現のみに依存しているため、PDTB の談話関係から分かることは「どの接続表現が使われているか」「(談話片対に接続表現が存在しない場合) どの接続表現が補完可能か」といった情報に留まり、文章内の出来事間の時間的關係を捉える上で談話関係が果たす役割を分析するには不十分である。また、文節談話表示理論 (Segmented Discourse Representation Theory ; SDRT)[4] は、談話関係がもたらす時間的制約や因果関係に対する制約などを形式的に定義しており、談話構造の計算機構を提供している。作例における典型的な場合、およびそれに類似する実例についてはよく説明できる一方、複数の関係が候補として考えられるような実例では判断が難しい場合が多くそのまま利用することはできない、という問題がある。

時間関係との依存関係を反映した談話関係の理論やアノテーションスキーマの確立が困難である理由の1つは、既存の談話関係において、解釈が揺れる「意図に関する関係」と、「明示的な情報から客観的に判断可能な関係」を区別せずに定義されている点にある。前者の関係の例として (7) を、後者の関係の例として (8) を見てみよう。

(7) a. 毎日ブドウにありがとう、と 言って、

b. 立派なブドウが 育った。

(8) a. ドアから部屋へ 突入した。

b. その部屋は 真っ暗だった。

(7) の「毎日ブドウにありがとう、と 言って」と「立派なブドウが 育った」という2つの節の間に因果関係はあるだろうか。人によっては「ない」と答えるだろうし、人によっては「ある」と答えるだろう。その判断は、読み手の信念に委ねられている。それに対して、(8) から「(発話者が) ドアから部屋へ突入したとき、その部屋の中は真っ暗だった」という状況が読み取れる、ということについて大多数の人は同意するはずである。また、1文目の述部が動作を表現する「突入した」であり、2文目の述部が状態を表現する「真っ暗だった」であること、いずれの述部も過去形であ

ることが判断の手掛かりとなっている、ということについても納得する人は多いと思われる。(7)のような「意図に関する関係」と、(8)のような「明示的な情報から客観的に判断可能な関係」が混在していると、後者の関係を捉える上で有効な情報と、談話関係との間の依存関係の分析を難しくすると考えられる。その理由は、エラー分析の際、そのエラーが読み手の信念から来る揺れなのか、手掛かり情報の不足による揺れなのかを判断できることが望ましいが、アノテーション結果のみからその判断を行うのは困難であるからである。

上記を踏まえると、談話理論およびそれに基づくアノテーションスキーマに望まれることは以下の2点である。

- 解釈が揺れる「話し手や書き手の意図に関する関係」を排除し、「出来事間の時間的・空間的な関係」に焦点を当て、明示的な情報により解釈が一意に定まる関係のみに絞って談話関係を再定義する。
- 定義した談話関係について、時間関係との依存関係を反映した、談話関係の決定手順を確立する。

上記の実現のために、本研究ではまず、出来事(eventuality)、モダリティ(modality)、時制(tense)、時間関係(temporal relation)と、談話関係との依存関係の整理を行い、それを元に談話関係の分類体系と客観的な談話関係の決定手順を定義する。その後、定義した分類体系と決定手順に従ってアノテーション実験を行い、一致率の算出やエラー分析を行うことにより、「解釈が一意に定まる関係のみに絞って談話関係を定義できたか」を検証し、談話関係の決定プロセスのどの部分を明らかにできたか、および、どの部分に課題が残るかを議論する。

本論文の構成は以下の通りである。2章では、談話関係の理論やアノテーションの関連研究について述べる。3章では、本研究で提案する談話関係アノテーションの具体的手順について説明する。4章では、アノテーション実験の方針について述べ、提案手法により構築されたコーパスの評価結果とエラー分析結果について述べる。5章では、本研究の結論と今後の課題について述べる。

第 2 章

関連研究

本章ではまず、談話関係の理論と談話関係が付与されたコーパス (corpus) の、代表的な研究について説明する。その後、関連研究の課題についてまとめる。

2.1 修辞構造理論 (RST)

談話関係の代表的な理論の 1 つとして、修辞構造理論 (Rhetorical Structure Theory ; RST)[1] がある。この理論では、談話構造は以下のように階層的な木構造として定義されている。

- (9) The Bush Administration, trying to blunt growing demands from Western Europe for a relaxation of controls on exports to the Soviet bloc, is questioning whether Italy's Ing C. Olivetti and company supplied military valuable technology to the Soviets.

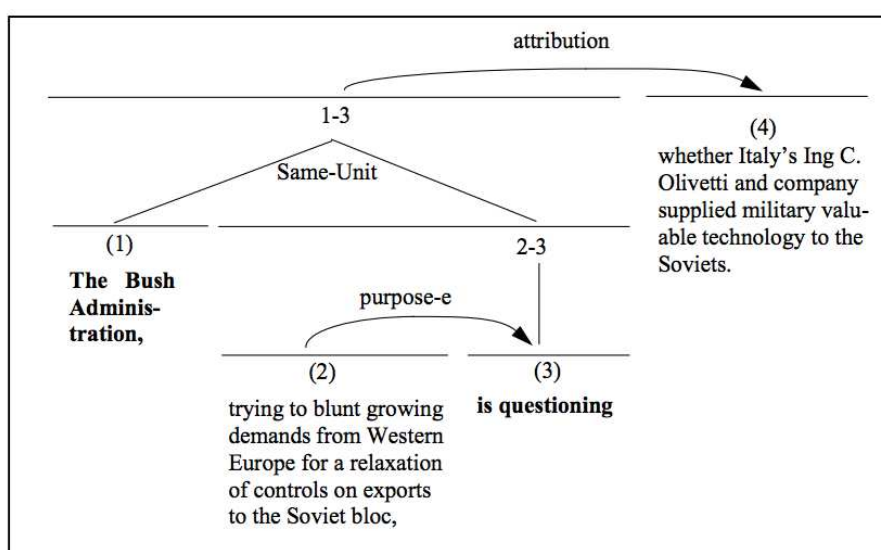


図 2.1 RST の談話構造の例 (Carlson et al.[5] p.25 Figure 3 より)

修辞構造理論の談話関係を付与したコーパスも実際に構築されており [5]、その規模は 385 文献、176,000 単語である。コーパスのテキストは アメリカ英語で書かれた Wall Street Journal の記事で、Penn Treebank[7] から選ばれている。

この理論では、テキスト上ではある連続する範囲 (Span) を談話関係を認定する対象とみなし、この範囲間の談話関係を考える。ここで、最小の Span は節であり、Elementary Discourse Unit (EDU) と呼ばれる。修辞構造理論では、EDU を階層的に構造化することでテキスト全体の談話構造を構築する。談話関係が付与された 2 つの Span は、典型的には一方が中心 (nucleus)、他方が周辺 (satellites) と呼ばれる。中心は、その談話関係の主となる役割を果たし、周辺は中心を補助する役割がある。中心と周辺の関係は、RST Discourse Treebank (RST-DT)[5] の時点では 57 種類の関係*¹が定義されており、これらの関係は、何らかのタイプの修辭的意味を共有する 16 のクラスに分けることができる (表 2.2)。各談話関係の定義は原則、「中心に関する制約」「周辺に関する制約」「中心と周辺に関する制約」「談話関係の効果」といった項目から構成される。例として、以下に RST における談話関係の定義の例を示す。

談話関係	定義
EVIDENCE	In an EVIDENCE relation, the situation presented in the satellite provides evidence or justification for the situation presented in the nucleus. Usually EVIDENCE relations pertain to actions and situations that are independent of the will of an animate agent. Evidence is data on which judgment of a conclusion may be based, and is presented by the writer or an agent in the article to convince the reader of a point. An evidence satellite increases the chance of the reader accepting the information presented in the nucleus.
ENABLEMENT	In an ENABLEMENT relation, the situation presented in the nucleus is unrealized. The action presented in the satellite increases the chances of the situation in the nucleus being realized.

表 2.1 RST-DT[5] における談話関係の定義の例

表 2.1 の EVIDENCE 関係の定義を見てみると、何が evidence (証拠) や justification (正当化) であるかについての判断基準に関して言及されていない。このように、修辞構造理論における談話関係は概念を厳密に定義していないものが少なくないため、アノテーションの際、作業者に主観的な判断を多分に強いることとなる。故に、判断が揺れやすい、再現性が担保されない、アノテータの判断の過程や根拠を結果から遡って分析することが難しい、等の問題がある。

*¹ Reference Manual [5] の 4.1 節では「53 種類の mononuclear 関係 (単一の EDU 間の関係) と、25 種類の multiclear 関係 (複数の EDU 間の関係) が用いられた」との記載があるが、Appendix では同名の mononuclear 関係と multiclear 関係を同一項目とみなし、57 種類の関係について説明しているため、ここでは 57 種類と表記した。

クラス	関係	クラスの概要説明
Attribution	attribution, attribution-n	表現活動を表す談話と、その内容との関係。
Background	background, circumstance	中心が解釈される文脈や背景を確立する。
Cause	cause, cause-resut, result	因果関係を表す。
Comparison	analogy, comparison, consequence, consequence-n, consequence-s, proportion	類推、比較などに関する関係。
Condition	condition, contingency, hypothetical, otherwise	条件、仮定などに関する関係。
Contrast	antithesis, concession, contrast	対比や譲歩に関する関係。
Elaboration	definition, elaboration-additional, elaboration-general-specific, elaboration-part-whole, elaboration-process-step, elaboration-object-attribute, elaboration-set-member	周辺が、中心に関する追加情報または詳細を提供する。
Enablement	enablement, purpose	未実現の目的と、それを実現するための行動を述べるような関係。
Evaluation	comment, evaluaton, evaluation-n, evaluation-s, example, interpretation, interpretation-n, interpretation-s	一方の談話が、もう一方の談話を評価したり解釈したりする。
Explanation	evidence, explanation-argumentative, reason	周辺が中心の理由、証拠を提供する。
Joint	disjunction, list	要素や選択肢を列挙する。
Manner-Means	manner, means	手段や方法を説明する。
Topic-Comment	comment-topic, problem-solution, problem-solution-n, problem-solution-s, question-answer, question-answer-n, question-answer-s, rhetorical-question, statement-response, statement-response-n, statement-response-s, topic-comment	質問と応答など、一方の談話が導入した一般的な話題を、もう一方で掘り下げる。
Summary	restatement, summary-n, summary-s	情報の要約、言い換えに関する関係。
Temporal	invertedsequence, sequence, temporal-before, temporal-after, temporal-same-time	時間的な関係。中心と周辺の内容のどちらが先に成立するか、または同時に成立するかを示す。
Topic Change	topic-shift, topic-drift	話題が転換した際に付与される。
Schemata	same-unit, textual-organization(author, abstract, column-title, date, footnote, heading, point-of-origin, sectiontext, section-title, source, text, title), span	タイトル、日付など、テキストの構造的要素を参照する。

表 2.2 RST-DT[5] における関係一覧（談話関係と Schemata の関係）

2.2 Discourse GraphBank (DGB)

Discourse GraphBank (DGB)[2] は、グラフ理論に基づく談話構造に、RST から派生した談話関係タグを付与した英語のコーパスである。グラフ構造を採用することにより、以下の例のように、木構造では記述できない、談話間の関係が交差する可能性も考えることができる。

- (10) a. There is a Eurocity train on Platform 1.
- b. Its destination is Rome.
- c. There is another Eurocity on Platform 2.
- d. Its destination is Zürich.

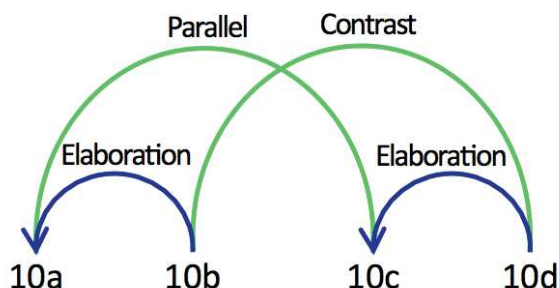


図 2.2 DGB の談話構造の例 (Wolf et al.[6] p.5 Figure 1 より)

また、RST では 1 つの節から他の複数の節への関係の付与ができなかったため、複数の関係として解釈可能な場合でも 1 つの関係のみを選択しなければならなかったが、DGB ではグラフ構造の採用により、1 つの節から他の複数の節への関係を考えることができるため、網羅的に関係ラベルの付与が行えるというメリットがある。一方、より多くの解釈の可能性を考慮する必要があるため、アノテーションの作業負担が大きいというデメリットがある。

DGB における談話関係ラベル一覧を表 2.3 に示す。ここで、関係タイプの *symmetrical* は、2 つの談話単位が対等であるような関係を意味し、*asymmetrical* は、一方の談話単位がより重要な役割を果たすような関係を意味する。表 2.3 に見られるように、DGB の談話関係もまた、客観的な付与基準が定義されておらず、読み手の信念に判断が委ねられている。故に、RST と同様に、再現性が担保されない、アノテータの判断の過程や根拠を結果から遡って分析することが難しい、などの問題がある。

なお、コーパスの規模は 135 文献、73,000 単語であり、RST-DT や後述の Penn Discourse Treebank よりも小さい。

関係ラベル名	関係タイプ	ラベルの定義
Parallel	symmetrical	Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then infer commonalities between members of E(DS0) and E(DS1) (This is the relation between two single DSs).
Parallel-B	symmetrical	Two groups of DSs are in a parallel relation.
Contrast	symmetrical	Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then infer contrasts between members of E(DS0) and E(DS1) (This is the relation between two single DSs). <ul style="list-style-type: none"> ・ Contrast-1 is a contrast between corresponding predicates in DS0 and DS1. The arguments of these contrasting predicates are identical. ・ Contrast-2 is a contrast between the arguments of corresponding predicates in DS0 and DS1. The predicates over these contrasting arguments are identical.
Contrast-B	symmetrical	Two groups of DSs are in a contrast relation.
Example	asymmetrical	<ul style="list-style-type: none"> ・ Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then find some element in E(DS1) that is a member or subset of the corresponding element in E(DS0). ・ Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then find some element in E(DS1) that is a new instantiation of an entity in E(DS0).
Generalization	asymmetrical	<ul style="list-style-type: none"> ・ Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then find some element in E(DS0) that is a member or subset of the corresponding element in E(DS1). ・ Infer a set of entities from DS0, E(DS0), and a set of entities from DS1, E(DS1). Then find some element in E(DS0) that is a new instantiation of an entity in E(DS1).
Elaboration	asymmetrical	Infer a set of coherent entities, E(DS0, DS1) from DS0 and DS1. The members of E(DS0, DS1) are centered around a common event or entity, e01.
Explanation	asymmetrical	Infer a causal relation between DS0 and DS1.
Violated Expectation	asymmetrical	Infer that normally there is a causal relation between DS0 and DS1 but that causal relation is absent between DS0 and DS1.
Condition	asymmetrical	the event described in the Nucleus can only take place if the event described in the Satellite also takes place (before or simultaneously with the event described in the Nucleus).
Temporal Sequence	asymmetrical	Infer a temporal sequence of the events described by DS0 and DS1. There is no causal relation between DS0 and DS1. If there is a causal relation, the relation between DS0 and DS1 should be described as a Cause-Effect relation.
Attribution	asymmetrical	The Satellite attributes the Nucleus to a source.
Same	symmetrical	a DS has intervening material; the “Same” relation is no coherence relation, but a “trick” that allows dealing with DSs nested in other DSs.

表 2.3 DGB[6] における関係一覧

2.3 Penn Discourse Treebank (PDTB)

Penn Discourse Treebank (PDTB)[3] もまた、RST から派生した談話関係を付与した、英語のコーパスの 1 つである。PDTB は、その名の通り Penn Treebank (PTB)[7] という、Wall Street Journal のニュース記事や Brown Corpus[8] (約 500 万語) に品詞情報や構文情報を付与したコーパスに対してアノテーションを施したもので、version 2.0 [9]、3.0 [10] の 2 種類が公開されている*2。コーパスの規模は、version 2.0 で 2,159 文献であり、既存の談話関係コーパスとしては最大である。このため、談話関係認識の研究で最も多く用いられている。

PDTB では、RST や DGB とは異なり、談話構造の形を仮定せず、隣接する談話間 (接続詞により繋がれた節間、同一段落内の隣接文間、コロン (:) またはセミコロン (;) で区切られた節と節間) の 2 項関係として表現されている。この手法では、談話間に接続表現が存在する場合は、その接続表現を手がかりとして談話関係の分類を行い、存在しない場合は、談話間に接続表現が挿入可能かどうか、どの接続表現が挿入できるか、の判断を行い、その結果を元に談話関係を分類している。以下に例を示す。

(11) a. While the earnings picture confuses,

b. observers say *the major forces expected to shape the industry in the coming year are clearer.*
(Prasad et al.[10] p.25 より)

(12) a. *Some Japanese operations, such as securities-trading rooms, may be ahead of their American counterparts,*

b. he says, but (Implicit=otherwise) “**basically, there’s little analysis done on computers in Japan.**” (Prasad et al.[10] p.27 より)

ここで、(11) は接続詞が存在する場合で、接続詞 “while” により Contrast (対比) という談話関係であると判断される例である。(12) は接続詞が存在しない場合で、接続詞 “otherwise” が挿入可能であることから、Exception (例外) という談話関係であると判断される例である。

この手法の貢献は、接続表現、ないしは接続詞挿入テストに基づいた、RST や DGB より客観的な判断基準を定義している点にある。また、表 2.4 のように 3 段階の階層的な関係分類となっており、目的に合わせて粒度の粗い関係のみを用いたり、全関係ラベルを用いたり、柔軟に利用できるという点も特徴的である。

*2 <https://www.seas.upenn.edu/~pdtb/>

Level-1	Level-2	Level-3
temporal	synchronous	—
	asynchronous	precedence succession
contingency	cause	reason
		result
		negresult
	cause+belief	reason+belief
		result+belief
	cause+speechact	reason+speechact
		result+speechact
	condition	arg1-as-cond
		arg2-as-cond
	condition+speechact	—
negative-condition	arg1-as-negcond	
	arg2-as-negcond	
negative-condition+speechact	—	
purpose	arg1-as-goal	
	arg2-as-goal	
comparison	concession	arg1-as-denier
		arg2-as-denier
	concession+speechact	arg2-as-denier+speechact
	contrast	—
similarity	—	
expansion	conjunction	—
	disjunction	—
	equivalence	—
	exception	arg1-as-excpt
		arg2-as-excpt
	instantiation	arg1-as-instance
		arg2-as-instance
	level-of-detail	arg1-as-detail
		arg2-as-detail
	manner	arg1-as-manner
arg2-as-manner		
substitution	arg1-as-subst	
	arg2-as-subst	

表 2.4 PDTB-3.0[10] の関係ラベルの階層関係

関係ラベル	概要説明
temporal	時間的な関係。
temporal.synchronous	時間的な重なりがある。
temporal.asynchronous	時間的な重なりがない。
temporal.asynchronous.precedence	Arg1 が Arg2 に先行する。
temporal.asynchronous.succession	Arg2 が Arg1 に先行する (precedence の逆)。
contingency	一方の引数が他方の引数の理由、説明、正当性を示すような関係。
contingency.cause	一方の引数が他方の引数から因果的に影響を受けるが、条件付きの関係にない。
contingency.cause.reason	Arg1 が理由、説明、正当性を示し、Arg2 が結果を示す。
contingency.cause.result	Arg2 が理由、説明、正当性を示し、Arg1 が結果を示す (contingency.cause.reason の逆)。
contingency.cause.negresult	Arg1 が、Arg2 に書かれた結果にならない理由、説明、正当性を示す。
contingency.cause+belief	聞き手に主張を信じさせるための根拠が提示される。
contingency.cause+belief.reason	Arg2 が、Arg1 で与えられた主張を正当化する証拠を提示する。
contingency.cause+belief.result	Arg1 が、Arg2 で与えられた主張を正当化する証拠を提示する (contingency.cause+belief.reason の逆)。
contingency.cause+speechact	speech act を発話している話者により、(発話した)理由が提示される。
contingency.cause+speechact.reason	Arg1 で発話者が speech act を発話する理由が Arg2 である。
contingency.cause+speechact.result	Arg2 で発話者が speech act を発話する理由が Arg1 である (contingency.cause+speechact.result の逆)。
contingency.condition	一方の引数 (条件) がまだ実現されていない状況を記述し、それが実現すると他方の引数 (結果) で記述された結果になる関係。
contingency.condition.arg1-as-cond	Arg1 が条件を、Arg2 が結果を表す。
contingency.condition.arg2-as-cond	Arg2 が条件を、Arg1 が結果を表す (contingency.condition.arg1-as-cond の逆)。
contingency.condition+speechact	結果が暗黙の speech act である condition (arg1-as-cond) が出現しなかったため、Arg2 が条件を、Arg1 が結果を表し、Arg1 の結果が暗黙の speech act であるものに付与されている。
contingency.negativecondition	一方の引数 (条件) がまだ実現されていない状況を記述し、それが実現しないなら、他方の引数 (結果) で記述された結果になる関係。
contingency.negativecondition.arg1-as-negcond	Arg1 が条件を、Arg2 が結果を表す。
contingency.negativecondition.arg2-as-negcond	Arg2 が条件を、Arg1 が結果を表す (contingency.negativecondition.arg1-as-negcond の逆)。
contingency.negativecondition+speechact	結果が暗黙の speech act である negativecondition。
contingency.purpose	一方の引数が、他方の引数によって達成される目的 goal のために、エージェントが実行する行動を表現する。
contingency.purpose.arg1-as-goal	Arg1 が目標を記述し、Arg2 がそれを達成するために取られた行動を記述する。
contingency.purpose.arg2-as-goal	Arg2 が目標を記述し、Arg1 がそれを達成するために取られた行動を記述する (contingency.purpose.arg1-as-goal の逆)。

表 2.5 PDTB-3.0[10] の関係ラベル一覧

関係ラベル	概要説明
comparison	2つの引数間の談話関係が、相違点または類似点（予想される結果と実際の結果との間の相違を含む）を強調する。
comparison.concession	想定される因果関係が、引数の1つで記述されている状況によって取り消されるか拒否される。
comparison.concession.arg1-as-denier	Arg2は何らかの予想される結果で、Arg1はそれを否定する。
comparison.concession.arg2-as-denier	Arg1は何らかの予想される結果で、Arg2はそれを否定する（comparison.concession.arg1-as-denierの逆）。
comparison.concession+speechact	1つの引数に関連付けられているspeech actが他の引数またはそのspeech actによって取り消されるか拒否される。
comparison.concession+speechact.arg2-as-denier+speechact	Arg2またはその関連するspeech actが、Arg1に関連するspeech actを取り消すか、拒否する。
comparison.contrast	Arg1とArg2の間の少なくとも2つの違いが強調表示されている。
comparison.similarity	Arg1とArg2の間の1つ以上の類似点により、各引数が全体として叙述していること、または、それが言及しているいくつかのエンティティに関して強調されている。
expansion	談話を拡張し、その物語や説明を前進させる関係。
expansion.conjunction	両方の引数が談話において生じた他の状況と同じ関係を持つときに使われ、2引数がその状況に対し同じ貢献をするか、一緒に貢献することを示す。
expansion.disjunction	2引数がどちらか一方または両方を保持した状態で選択肢として提示されるときに使われ、2引数がその状況に対し一緒に貢献することを示す。
expansion.equivalence	両方の引数が同じ状況を、異なる観点から記述するために使用される。
expansion.exception	一方の引数が、説明した状況が成り立つ一連の状況を引き起こすときに使用され、もう一方の引数は、そうでない1つ以上のインスタンスを示す。
expansion.exception.arg1-as-excpt	Arg1が例外を示す。
expansion.exception.arg2-as-excpt	Arg2が例外を示す（expansion.exception.arg2-as-excptの逆）。
expansion.instantiation	一方の引数が一連の状況の中で成立している状況を説明し、もう一方の引数がそれらの状況の1つ以上を説明するときに使用される。
expansion.instantiation.arg1-as-instance	Arg1がArg2により記述された状況の1つ以上のインスタンスを提供。
expansion.instantiation.arg2-as-instance	Arg2がArg1により記述された状況の1つ以上のインスタンスを提供（expansion.instantiation.arg1-as-instanceの逆）。
expansion.level-of-detail	両方の引数が同じ状況を記述しているが、詳細不明という場合に使用。
expansion.level-of-detail.arg1-as-detail	Arg1がArg2の状況をより詳細に記述するときに使用。
expansion.level-of-detail.arg2-as-detail	Arg2がArg1の状況をより詳細に記述するときに使用（expansion.level-of-detail.arg2-as-detailの逆）。
expansion.manner	ある引数で記述された状況が、他の引数で記述された状況が発生した、または行われた方法を表すときに使用。
expansion.manner.arg1-as-manner	mannerを表すのがArg1の場合。
expansion.manner.arg2-as-manner	mannerを表すのがArg2の場合（expansion.manner.arg2-as-mannerの逆）。
expansion.substitution	引数が排他的な選択肢として提示され、除外されている場合に使用される。
expansion.substitution.arg1-as-subst	Arg2に関連した状況が除外された後に残っている選択肢をArg1が伝えるときに使用。
expansion.substitution.arg2-as-subst	Arg1に関連した状況が除外された後に残っている選択肢をArg2が伝えるときに使用（expansion.substitution.arg1-as-substの逆）。

表 2.6 PDTB-3.0[10] の関係ラベル一覧（続き）

以下に PDTB-3.0 における談話関係の定義の例を示す。

談話関係	定義	接続詞の例
contingency.cause+belief	This tag is used when evidence is provided to cause the hearer to believe a claim. The belief is implicit.	as, because, by, from, given, in, indeed, so, thus, when, with
comparison.concession	This tag is used when an expected causal relation is cancelled or denied by the situation described in one of the arguments.	as, as if, as much as, but, but then again, but then, by, despite, even after, even as, even before then, even before, even if, even so, even then, even though, even when, even while, even with, for, however, if, if only, in any case, in fact, in the end, indeed, meanwhile, nevertheless, no matter, nonetheless, nor, not only+but, on the other hand, only, or, regardless, regardless of, still, though, whatever, when, whether, while, with, without, yet

表 2.7 PDTB-3.0[10] における談話関係の定義の例

定義に見られるように、接続表現に基づいている部分以外の判断基準は厳密に定義されておらず、接続詞挿入テストにおいて「どの接続詞を挿入するか？」の判断は主観に委ねられている。例えば、表 2.7 の contingency.cause+belief 関係の定義では、どのような場合に“evidence is provided to cause the hearer to believe a claim”といえるのかについての判断基準は与えられていない。このことは、談話関係の定義が接続詞（の挿入可能性）のみに依存していることを意味し、それ自体分析の確立していない接続詞という現象に談話関係という現象を還元していることになってしまう。接続詞に還元されたところでそこから得られる情報が少ないため、PDTB の談話ラベルが認識できることによって分かることは、どのような接続詞が挿入可能なのかに留まり、これは談話関係の重要な側面を見落としていることになる。故に、PDTB の談話関係が認識できたとしても、時間関係認識等のタスクで直接役立つには難しいと考えられる。

2.4 分節談話表示理論 (SDRT)

分節談話表示理論 (Segmented Discourse Representation Theory ; SDRT)[4] は、動的意味論の 1 つである談話表示理論 [11] と談話構造の説明を統合し、談話解釈のための意味論を提案している。この手法では、談話関係がもたらす時間的制約や因果関係に対する制約などを形式的に定義しており、談話構造の計算機構を提供している。以下に、SDRT における時間と因果に対する制約の定義の例を示す。

Cause and Subtype aren't Occasion (Asher et al.[4] p.207 より)

(a) $cause_D(\sigma, \beta, \alpha) \rightarrow \neg occasion(\alpha, \beta)$

(b) $subtype_D(\sigma, \beta, \alpha) \rightarrow \neg occasion(\alpha, \beta)$

ここで、 $cause_D(\sigma, \beta, \alpha)$ は、談話 σ において、 β は α の原因である、ということの意味し、 $occasion(\alpha, \beta)$ は、 α が起きた後に β が起こる、ということの意味している。また、 $subtype_D(\sigma, \beta, \alpha)$ は、談話 σ において、 β は α のサブタイプである、ということの意味している。つまり、(a) は「 β が α の原因であるならば、 α が β より先に起こることはない」ということ、(b) は「 β が α のサブタイプであるならば、 α が β より先に起こることはない」ということを示している。各談話関係における形式的定義、時間的制約や因果関係に対する制約も同様に定義されている (表 2.8, 表 2.9)。表 2.8, 表 2.9 に見られるように、RST、DGB、PDTB と比べ、談話関係と時間関係、因果関係の対応はより詳細に定義されているものの、以下のように複数の関係のどれに該当するかで判断に悩む例が少なくなく、そのまま実テキストのアノテーションに利用することはできない、という問題がある。

(13) a. 眠気が消えて、

b. やる気が出てきた。

(14) a. ストーブが点いていた。

b. いつの間にか、猫が焦げていた。

ここで、(13) は Narration と Result とで揺れる可能性がある例である。「消える」と「出る」の間に因果関係がある、と判断すると Result となり、因果関係がない、と判断すると Narration となる。SDRT では因果関係の有無を判断するための基準までは提供されていないため、その判断は読み手の信念に委ねられることになる。また、(14) は Narration と Background で揺れる可能性のある例

である。Narration と Background とで揺れるのは、「点く」という出来事が起きてから「焦げる」という出来事が起こる、と判断すると $\phi_{Narration(a,b)} \Rightarrow overlap(prestate(e_b), Adv_b(poststate(e_a)))$ という時間的制約によって Narration となり、「点いている」という状態と「焦げている」という状態に時間的重なりがある、と判断すると $\phi_{Background(a,b)} \Rightarrow overlap(e_a, e_b)$ という時間的制約によって Background となるためである。SDRT の時間的制約の定義では、出来事が動作 (event) であるか、状態 (state) であるかの区別をしていないため、「元は動作を表す動詞であるが、テイルが後接することで状態となった述部」間などでの関係の判断基準が曖昧になってしまっている。

SDRT は談話片対の決定方法も特徴的である。PDTB 等と異なり、SDRT では、隣接していない談話セグメントも接続候補となり得るような定義になっている。以下にその定義を示す。

Right Frontier Constraint (RFC) (Asher et al.[4] p.148 Definition 14 より)

Suppose that a constituent β is to be attached to a constituent in the SDRS with a discourse relation other than Parallel or Contrast. Then the available attachment points for β are :

1. The label $\alpha = \text{LAST}$;

2. Any label γ such that:

(a) i-outscopes(γ, α) (i.e. $R(\delta, \alpha)$ or $R(\alpha, \delta)$ is a conjunct in $F(\gamma)$ for some R and some δ); or

(b) $R(\gamma, \alpha)$ is a conjunct in $F(\lambda)$ for some label λ , where R is a subordinating discourse relation.

We gloss this as $\alpha < \gamma$.

3. Transitive Closure:

Any label γ that dominates α through a sequence of labels $\gamma_1, \gamma_2, \dots, \gamma_n$ such that $\alpha < \gamma_1 < \gamma_2 < \dots < \gamma_n < \gamma$

上記の定義は「最後に談話が接続したノード (LAST) と、LAST より上にあるノードのうち、右側に coordinating relation が接続していないものが接続候補先となりうる」ということを意味している。ここで、coordinating relations は、水平方向に接続する関係で、subordinating relations は垂直方向に接続する関係である。次の例で、RFC の振る舞いを確認してみよう。

(π_1) John had a great evening.

(π_2) He had a great meal.

(π_3) He ate salmon.

(π_4) He devoured cheese.

(π_5) He then won a dance competition. (Asher et al.[4] p.139 より)

この例を、SDRT の意味表示 SDRS に変換すると以下のようなになる。

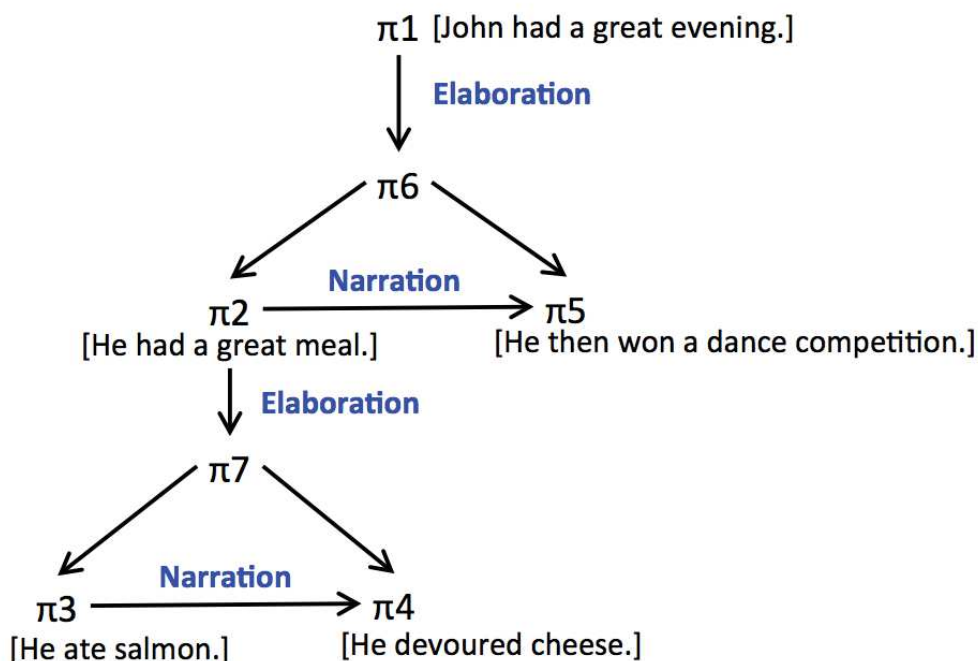


図 2.3 (π1)-(π5) の意味表示 (Asher et al.[4] p.147 Figure 4.5 より)

この例に次の文が後接する場合、接続候補先となるノードは LAST に該当する π5、「LAST より上にあるノードのうち、右側に coordinating relation が接続していないもの」に該当する π1、π6^{*3}である。SDRT では、このように談話が接続可能となる条件を設けることで探索空間を小さくし、「接続先を隣接する談話に限定しない」ことにより接続可能な談話の数が劇的に増加してしまうという問題を回避している。しかしながら、RFC を導入することによって、制約に違反する談話間には談話関係を付与できないという問題や、アノテータ間で、異なる接続先を選択してしまった場合、および、接続先は同じでも水平・垂直が異なるラベルを付与してしまった場合、その後の談話構造やそれらに付与されるラベルが大きく異なってしまうといった問題が生じているという側面もある。

^{*3} π6 は、π2 と π5 を含む談話の集合。

談話関係	定義
Continuation	<p>Just like <i>Narration</i> save it lacks its spatio-temporal consequences :</p> <ul style="list-style-type: none"> • It's veridical (and so satisfies the Satisfaction Schema). • <i>Continuation</i> is subject to the same topic constraint as <i>Narration</i>.
Narration	<p>Informally, this relation holds if the constituents express eventualities that occur in the sequence in which they were described. It can connect indicatives or requests. More formally :</p> <ul style="list-style-type: none"> • It's veridical, and so it satisfies the Satisfaction Schema. • Topic Constraint on Narration : $\phi_{Narration(\alpha,\beta)} \Rightarrow \neg\Box(K_\alpha \sqcap K_\beta)$ I.e., α and β share a contingent, common topic (and the more informative the topic, the better the narration). • Spatiotemporal Consequence of Narration : $\phi_{Narration(\alpha,\beta)} \Rightarrow \text{overlap}(\text{prestate}(e_\beta), \text{Adv}_\beta(\text{poststate}(e_\alpha)))$ I.e., where things are in space and time at the end of e_α is where they are at the beginning of e_β.
Result	<p><i>Result</i> connects a cause to its effect :</p> <ul style="list-style-type: none"> • It's veridical, and so it satisfies the Satisfaction Schema. • Axiom on Result : $\phi_{Result(\alpha,\beta)} \Rightarrow \text{cause}(e_\alpha, e_\beta)$
Contrast	<ul style="list-style-type: none"> • <i>Contrast</i> is veridical, and so it satisfies the Satisfaction Schema. • K_α and K_β must have similar <i>semanticstructures</i>. That is, there is a partial isomorphism between the DRS-structure of K_α and that of K_β. All else being equal, the closer the mapping is to an isomorphism, the better the <i>Contrast</i> relation. • There must be a contrasting theme between K_α and K_β. This is computed on the basis of the above partial isomorphism between the structures K_α and K_β. All else being equal, the more contrasting the theme, the better the <i>Contrast</i> relation. Degree of contrast is defined the themes of degree of difference between the propositions which mark the themes on the nodes of the semantic structure. The maximal constituent negates a default consequence of the other.
Parallel	<ul style="list-style-type: none"> • <i>Parallel</i> is veridical, and so it satisfies the Satisfaction Schema. • K_α and K_β must have similar <i>semanticstructures</i>. • There must be a common theme between K_α and K_β. <p>This is computed on the basis of the above partial isomorphism between the structures K_α and K_β. The more informative the common theme, the better the <i>Parallel</i> relation.</p>
Precondition	<p><i>Precondition</i>(α, β) is used to represent “anti-narration” (see below). That is, <i>Precondition</i> is like <i>Narration</i> except that the order of the arguments is reversed.</p>
Alternation	<p>This is the rhetorical relation equivalent to <i>or</i> (or more technically, dynamic \vee) :</p> <ul style="list-style-type: none"> • $(w, f)[[Alternation(\alpha, \beta)]]_M(w', g)$ iff $(w, f)[[K_\alpha \vee K_\beta]]_M(w', g)$ (and hence $(w, f) = (w', g)$)
Consequence	<p>This corresponds to dynamic \Rightarrow :</p> <ul style="list-style-type: none"> • $(w, f)[[Consequence(\alpha, \beta)]]_M(w', g)$ iff $(w, f)[[K_\alpha \Rightarrow K_\beta]]_M(w', g)$ (and hence $(w, f) = (w', g)$)

表 2.8 SDRT[4] における談話関係一覧 (Coordinating Relations)

談話関係	定義
Background	<p>This relation holds whenever one constituent provides information about the surrounding state of affairs in which the eventuality mentioned in the other constituent occurred.</p> <ul style="list-style-type: none"> • It's veridical, and so satisfies the Satisfaction Schema. • Temporal Consequence of Background : $\phi_{Background(\alpha,\beta)} \Rightarrow overlap(e_\beta, e_\alpha)$ • If the SDRS contains $\pi' : Background(\pi_2, \pi_1)$, then it also contains $\pi : K_\pi$ 'repeats' the contents of K_{π_1} and K_{π_2} and $\pi'' : FBP(\pi, \pi')$
Elaboration	<p>This is a <i>subordinating</i> relation :</p> <ul style="list-style-type: none"> • It's veridical, but it satisfies a slightly modified version of the Satisfaction Schema. • $Elaboration(\alpha, \beta) \vdash \downarrow (\alpha, \beta)$ • Temporal Consequence of Elaboration : $\phi_{Elaboration(\alpha,\beta)} \Rightarrow Part-of(e_\beta, e_\alpha)$
Explanation	<p>This is a <i>subordinating</i> relation, and it's the 'dual' to <i>Result</i>:</p> <ul style="list-style-type: none"> • It's veridical, and so satisfies the Satisfaction Schema. • Temporal Consequence of Explanation : <p>(a) $\phi_{Explanation(\alpha,\beta)} \Rightarrow (\neg e_\alpha \prec e_\beta)$ (b) $\phi_{Explanation(\alpha,\beta)} \Rightarrow (event(e_\beta) \Rightarrow e_\beta \prec e_\alpha)$</p>
Commentary	<p>$Commentary(\alpha, \beta)$ holds if β provides an opinion or evaluation of the content associated with α. "Supplementary adverbs" are good surface cues for Commentary.</p>

表 2.9 SDRT[4] における談話関係一覧 (Subordinating Relations)

2.5 日本語における談話関係アノテーションの研究

日本語における談話関係アノテーションの研究として河原らのグループものがある [13, 14, 15]。大規模な談話関係アノテーションを行う際に長い時間と高いコストがかかるという問題を解決するために、この研究ではクラウドソーシングを用いて談話関係コーパスを構築している。クラウドソーシングでは、専門家によるアノテーションと比較して品質が劣るという欠点がある。したがって、品質の低下を防ぐために、10 人に対して同じ設問を実施し、重み付き投票を行っている。また、これまでの典型的な談話関係アノテーションをそのままクラウドソーシングで行うことは困難であるため、談話関係アノテーションの手続きを簡略化している。具体的には、「節ペアに対する談話関係の有無の判定」と「談話関係があると判定された節ペアに対する談話関係タイプの判定」の 2 段階にタスクを分割している。1 段階目では、図 2.4 のように各文書に含まれる節ペアごとに談話関係を持つか否かの 2 値分類を行い、2 段階目では、図 2.5 のように、1 段階目で談話関係を持つ可能性があるると判定された節ペアに対して最も適切な談話関係を選択する。

【…】で示されるフレーズの間、論理的な関係（原因結果、条件、目的）があるかどうかを判断し、選択肢から選んでください。

【1 十五夜というとお月見です。】
 【2 お月見はお花見同様、日本人の誰もが知る年中行事の一つです。】
 【3 十五夜という、月見団子とススキです。】

1 と 2 の間に論理的な関係がある

1 と 3 の間に論理的な関係がある

2 と 3 の間に論理的な関係がある

どのフレーズの間にも論理的な関係はない

確定して次へ

[前の設問に戻る](#)

図 2.4 節ペアに対する談話関係の有無の判定のタスク画面（河原ら [15] p.820 図 1 より）

文章中の【…】で示されることばの間に成り立つ関係について、選択肢から選んでください。

郡市代表 16 チームが熱戦を展開します。

【当日は交通規制を行います。】
【警察官や係員の指示に従って】
ご協力をお願いします。

- 原因・理由
- 目的
- 条件
- その他根拠
- 対比
- 譲歩
- 上記のいずれの関係もない

例:

・原因・理由（【雨が降ったから】 ↔ 【道が濡れている】，【体調が悪いので】 ↔ 【遊びに行かなかった】）

図 2.5 談話関係があると判定された節ペアに対する談話関係タイプの判定のタスク画面（河原ら [15] p.820 図 2 より）

関係ラベルもまた、PDTB をベースとして表 2.10 のように単純化されている。

上位タイプ	下位タイプ	例文
根拠・条件	原因・理由	1. 花が陰暦 5 月に咲くため 2. 「皐月」と呼ばれている。
	目的	1. 「泉響」に耳を傾けるには、 2. やはり泊まるしかない。
	条件	1. あなたが本気ならば、 2. 真実を教えましょう。
	その他根拠	1. ここにカバンがあるから、 2. まだ社内にいるだろう。
転換	対比	1. 雪は降らないけれど、 2. コットンが舞う。
	譲歩	1. あのレストランは確かにおいしいが、 2. 値段は高い。
談話関係なし・その他弱い関係		—

表 2.10 河原らの談話関係一覧 [13, 14, 15]

この研究における談話関係は「根拠・条件」「転換」「その他」に限定されている。この研究では、談話関係認識の応用先として、言論間の関係を同定して、賛否両論あるトピックに対する俯瞰的なマップを生成する言論マップ [16] などへの活用を想定しており、この目的には詳細化や換言といった付加的な談話関係や時間関係は重要ではない、と河原らが判断したためである。故に、本研究では「出来事間の時間的・空間的な関係」に焦点を当てているが、河原らの研究では、解釈が揺れる「話し手や書き手の意図に関する関係」に主眼を置いている、という違いがある。

2.6 関連研究における課題と本研究の位置付け

本節では、2 章で述べてきた問題点を総括する。まず、各関連研究の問題点を以下にまとめる。

- RST と DGB における談話関係の定義には、概念を厳密に定義していないものがある。そのような概念間で判断を行う場合は、多分に主観的な判断というものを作業者に強いらざるを得ない。
- PDTB における談話関係は、談話関係とその判断基準の定義が接続表現（の挿入可能性）のみに依存しているため、PDTB の談話関係から分かることは「どの接続表現が挿入可能か」に留まる。

- また、PDTB における談話関係は、接続表現に基づいている部分以外の判断基準は厳密に定義されておらず、接続詞挿入テストにおいても「どの接続詞を挿入するか？」の判断は主観に委ねられている。
- SDRT における談話関係では、RST、DGB、PDTB と比べ、談話関係と時間関係、因果関係の対応がより詳細に定義されている。しかしながら、因果関係の有無を判断するための客観的な基準までは提供されておらず、因果関係の有無の判断が伴う談話関係については、読み手の信念に判断が委ねられてしまっている。
- また、SDRT の談話関係の時間的制約の定義では、出来事が動作 (event) であるか、状態 (state) であるかの区別をしていないために、時間関係の判断、および、その判断を伴う談話関係の判断基準が一部曖昧になってしまっている。

上記の問題点は以下の 2 点に集約される。

- 談話関係に関するいずれの関連研究も、談話関係と時間関係の依存関係を反映した設計となっているとは言い難い。
- また、談話関係に関するいずれの関連研究も、解釈が揺れる「多分に主観的な判断を伴う関係」と、明示的な情報により解釈が一意に定まる関係とを十分に選り分けられていない。

これらの問題が、談話関係と時間関係の決定プロセスの分析、ひいては、時間関係の依存関係を反映した談話関係の理論構築を阻んでいると推察される。したがって、本研究ではそのような主観的な判断を可能な限り排除することを目指す。主観的な判断というものに関しては次の点に注意しなければならない。(i) 定義されていない概念同士の間で、どちらがより対象に対する判断にふさわしいかを根拠なく判断するケースと、(ii) 文自体の意図あるいは意味が曖昧で、その判断に聞き手の世界知識や常識を利用しなければ解釈ができないケースがあり、この 2 つの場合は区別しなければならない、という点である。本研究で排除したいのは (i) のケースで、(ii) のケースには、例えば以下のような例が含まれる。

(15) パーティが行われ、バンド演奏があった。

この例では、パーティが始まってから演奏するので、開始時には明らかな前後関係がある。終了時について文では言及していないが、常識的に考えると、パーティが終わってからバンドが演奏されていることはないはずなので、「パーティを行う」という出来事が「バンド演奏がある」という出来事を時間的に包含するはずである。このような判断は、主観で行っていると言えなくもな

いが、談話関係の判断に本質的に含まれると考えられる。

上記の問題点や議論を踏まえ、本研究では、談話関係アノテーションの定義から基準のない判断を可能な限り減らすために、以下5点について取り組む。

- 因果関係の付与は、因果を表す接続表現が明示された談話間に留めるなど、文章内に明示的に現れる情報に基づいて、客観的に判断可能な談話関係の判断基準を定義する。
- 談話関係の判断に関わると考えられる明示的な情報（出来事、モダリティ、時制）のアノテーション基準を定義する。
- 出来事が動作 (event) であるか、状態 (state) であるかを考慮し、談話関係の時間的制約を定義する。
- 文章内に明示的に現れる情報（出来事、モダリティ、時制）、および、時間関係との依存関係を反映した、談話関係を設計する。
- 文章内に明示的に現れる情報（出来事、モダリティ、時制）と時間関係、談話関係の決定手順を定義する。

アノテーションスキーマの詳細については3章で、アノテーション実験とその結果の詳細については4章で説明していく。

第 3 章

談話関係アノテーション手法の提案

本章では、本研究で提案する談話アノテーションの定義について説明する。まず、3.1 節で、本研究の作業の全体像を概観する。次に、3.2 節では、談話関係の付与対象となる節の述部の判定方法について述べ、3.3 節では、談話関係の判定の際に手がかりとなる、出来事とモダリティの情報の付与方法について述べる。その後、3.4 節でどの述部とどの述部を接続するかについての判定方法について、および、3.5 節で談話関係の付与方法について説明し、3.6 節で提案手法の有効性を定性的に評価する。

3.1 作業概要

ここでは、本研究におけるアノテーション作業の全体像を概説する。本研究ではまず、文章中の談話関係の付与対象となる述部の同定を行い、活用形・時制ラベルと、出来事とモダリティの情報を付与する。その後、どの述部間で接続するかの判定を行い、談話関係ラベルの付与を行う。例えば、次の文章が与えられたとする。

22 日午前 3 時頃、山田太郎（仮名）さん方に外国人らしい男数人が押し入り、山田さんと妻の花子さんの顔や手足をテープで縛り、金品を奪って逃走。約 1 時間後、花子さんは自力でテープをほどいて隣家に駆け込み 110 番通報した。

図 3.1

まず、述部の同定と、活用形・時制ラベルと、出来事とモダリティの情報の付与を行うと、図 3.2 のようになる。「押し入る」「縛る」「奪う」「逃走する」「ほどく」「駆け込む」「通報する」は非状態動詞なので、出来事が動作を表しているという情報「event」が付与されている。また、「通報する」は過去の助動詞「た」が付いているため、「過去」という情報が付加された活用形ラベルが付与されている。

22日午前3時頃、山田太郎(仮名)さん方に外国人らしい男数人が押し入り、
連用形, event

山田さんと妻の花子さんの顔や手足をテープで縛り、
連用形, event

金品を奪って逃走。
テ形, event 省略形, event

約1時間後、花子さんは自力でテープをほどいて
テ形, event

隣家に駆け込み110番通報した。
連用形, event 終止形-過去, event

図 3.2 述部に活用形・時制ラベルと出来事の情報を与与した例

上で付与した情報を踏まえ、図 3.1 の文章に接続先判定と談話関係ラベルの付与までを行うと、図 3.3 のようになる。今回は、文内の場合は係り受け関係にある述部間を繋ぎ、文間の場合は隣接文間における主節の述部間を繋ぐとしたため、図の赤矢印のような接続関係となっている。

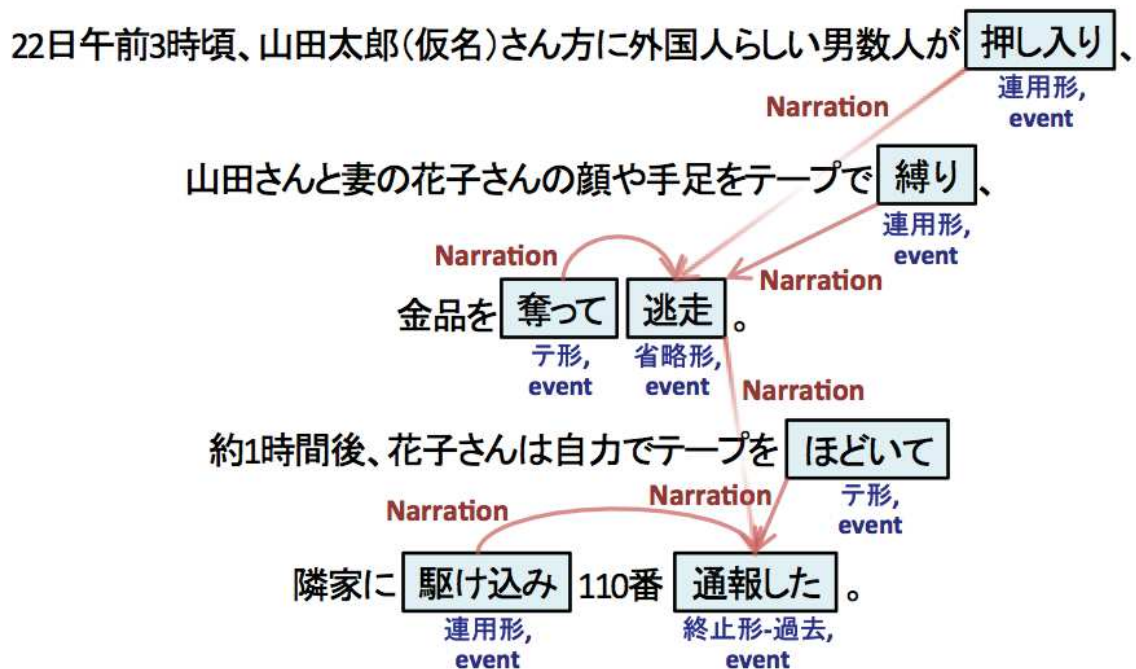


図 3.3 接続先判定と談話関係ラベルの付与を行った例

各関係における述部の出来事は、1つ目の event が先に起こり、その後2つ目の event が起こる、という順序になっていると解釈できるため、述べた順番通りに出来事が成立するような談話関係である Narration を付与している。

このように述部に付与された情報を踏まえつつ、接続先判定と談話関係ラベルの特定を行う。

3.2 節の述部のアノテーション

本研究では、節の述部間に談話関係をアノテーションする。故に、談話関係アノテーションを行うにあたり、まず節の述部に対して、活用形・時制ラベルの付与を行う。RST、DGB、PDTB のようにスパン (Span) を談話単位 (アノテーション対象) とすること、および、SDRT のように文を論理式 (意味表示) に変換したものの部分論理式を談話単位とすることには次のような問題点があるためである。

- スパンを談話単位とする場合、埋め込み等、例外となるケースが多く、各ケースの対応方法を定める負担や、人手で行う場合は作業負担が大きくなってしまう。また、(半)自動で行う場合は場合分けが膨大になってしまう。
- 文を論理式 (意味表示) に変換したものの部分論理式を談話単位とする場合、Boxer[17] や ccg2lambda[18] のような論理式を生成するツールの精度が (特に入力文が長い場合に) 十分でなく、手作業による修正が必要になるが、論理式に対する人手修正作業はコストが高いという問題がある。

スパンを談話単位とする場合に例外となる例を以下に示す。

(16) 太郎は、インフルエンザに罹ったので、病院に行った。

(17) 次郎は 119 番に、三郎は家族に、電話を掛けた。

(16) は「(太郎は) インフルエンザに罹ったので」と「太郎は病院に行った」を談話単位として指定したいが、主節「病院に行った」とその項「太郎は」が従属節「インフルエンザに罹ったので」により分断されてしまっている。また、従属節は主語「太郎は」を主節と共有しており、「太郎は」が主節の一部としてしまうと主語を持たないことになってしまう。一方、(17) は「次郎は 119 番に電話を掛けた」と「三郎は家族に電話を掛けた」を談話単位として指定したいが、「次郎は 119 番に」と「三郎は家族に」の項が「電話を掛けた」という述部を共有している。そのため、述部を共有している場合の対応方法を新たに決める必要がある。スパンを談話単位とする場合、このように例外となるケースが数多く存在するという問題点がある。

述部を談話単位とする理由は上述の問題点だけではない。以下のような利点があるというのも理由の1つである。

- 述部も、テキスト上ではある連続する範囲として現れるため、スパンを談話単位とする場合と同様にアノテーションツールでアノテーションしやすい。
- 時制節の述部は、意味論的に時区間を導入しているといえるので、時制節の述部間とすることで時区間同士の関係とみなすことができる。したがって、論理式を談話単位とする場合と同じく、意味論上の対応物が存在する。
- 述部のアノテーションは、品詞タグ付け、係り受け解析と同時に行えるタスクとなっており、言語理解プロセス全体の自動化を目指す場合、自然に組み込める見通しがある。

また、スパンベースの場合、動詞句の等位接続の場合は共有している項を除いた部分をスパンとみなすことになるが、それを突き詰めていくと結局述部ベースに収束すると考えられる。ここまで述べてきた問題点、利点等を踏まえ、本研究では述部を談話単位としてアノテーションを行う。

3.2.1 節の述部の判定

本研究では、与えられた節において、ある用言（動詞、形容詞、状詞）の語幹から始まり、かつ、図 3.4 の状態遷移図から生成される語の並びのうち、最長のものを**述部**と呼ぶ。

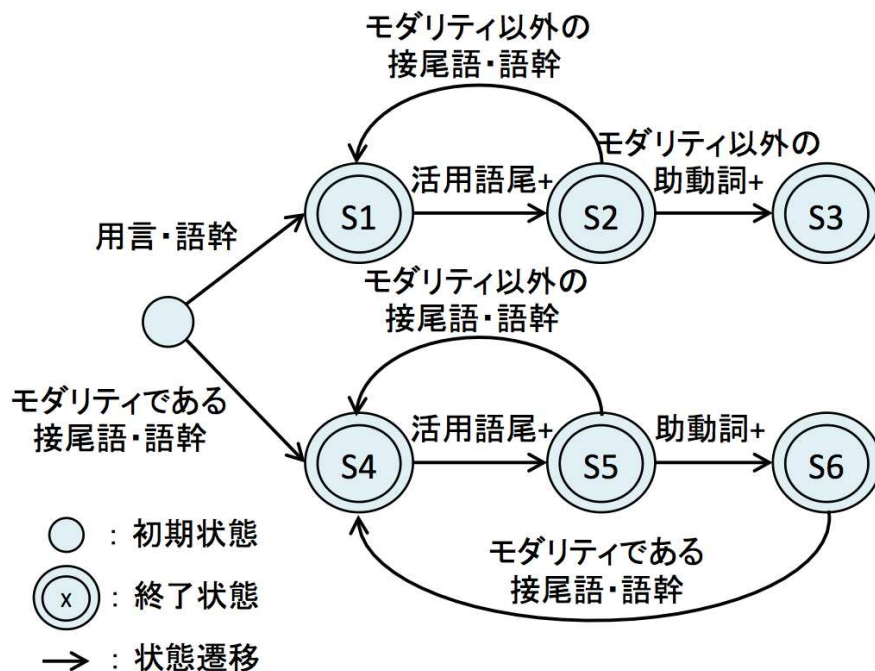


図 3.4 述部判定の状態遷移図

例えば、以下の(18)(19)(20)(21)では下線部が述部である。

- (18) 2018年4月に入学。
- (19) 太陽の方角へ走る。
- (20) さっきまで雨が降っていたはずだ。
- (21) まるで鉄の上に座っているようだったらしい。

各文の述部が上記の状態遷移図でどのように遷移するかを見てみると、次のようになる。

- (18) 「入学」：初期状態に動詞の語幹「入学」が与えられ、S1に遷移する。「入学」に後続する語はなく、S1は終了状態であるため、終了。
- (19) 「走る」：まず初期状態に動詞の語幹「走」が与えられ、S1に遷移。次に、S1に活用語尾「る」が与えられ、S2に遷移。「る」に後続する語はなく、S2は終了状態であるため、終了。
- (20) 「降っていた」：初期状態に動詞の語幹「降」が与えられ、S1に遷移。次に、S1に活用語尾「っ」「て」が与えられ、S2に遷移。S2に動詞性接尾語の語幹「(て)い(る)」が与えられ、S1に遷移し、空活用語尾によってS2に遷移し、助動詞「た」が与えられ、S3に遷移。「た」に後続する語はなく、S3は終了状態であるため、終了。
- (20) 「はずだ」：初期状態にモダリティである接尾語・語幹「はず(だ)」が与えられ、S4に遷移。次に、空活用語尾によってS5に遷移し、助動詞「だ」が与えられ、S6に遷移。「だ」に後続する語はなく、S6は終了状態であるため、終了。
- (21) 「座っている」：初期状態に動詞の語幹「座」が与えられ、S1に遷移。次に、S1に活用語尾「っ」「て」が与えられ、S2に遷移。S2に動詞性接尾語の語幹「(て)い(る)」が与えられ、S1に遷移し、その後活用語尾「る」が与えられ、S2に遷移。「る」に後続する語はなく、S2は終了状態であるため、終了。
- (21) 「ようだったらしい」：モダリティである接尾語・語幹「よう(だ)」が与えられ、S4に遷移。次に、助動詞「だっ」によってS5に遷移し、助動詞「た」によりS6に遷移。その後、モダリティである接尾語・語幹「らし(い)」が与えられ、S4に遷移し、活用語尾「い」によりS5に遷移。「い」に後続する語はなく、S5は終了状態であるため、終了。

用言の語幹・活用語尾・接尾語の語幹・助動詞の判断は、『日本語文法の形式理論』[19]に従う。詳細な判断は、3.3.1節の表3.1を参照。その他の注意点は以下の通りである。

- タイトルや見出しに含まれる述部はアノテーション対象外とする。

- 句読点、文末の記号（「！」など）、接続助詞、終助詞、係助詞など、図 3.4 に含まれないものは述部に含まないこととする。
- モダリティが複数出現する場合、(21)「ようだったらしい」のように一塊の述部として扱う。
- 「普通名詞+だ」「形式名詞+だ」の「普通名詞」「形式名詞」も「用言（状詞）の語幹」であることに注意する。例えば「太郎だ」であれば、「だ」ではなく「太郎だ」で述部とする。
- 「名詞+になる」「名詞+とする」の場合は「なる」「する」のみをアノテーション対象とする。「楽になる」のように状詞が含まれることもあるが、この場合も「なる」のみを付与対象とする。

3.2.2 モダリティの判定

モダリティとは、ある命題に対する書き手の認識や態度を表す言語表現である。モダリティに埋め込まれた出来事が成立する時区間と、モダリティが表す書き手の認識が成立する時区間が異なることがある。以下に例を示す。

(22) a. 財布をどこかに 落とした ようだ。

b. 鞆の中には 入っていない。

ここで、「落とした」という出来事が成立したのは、モダリティ「ようだ」が成立する時区間、つまり、この文の発話者が「落とした」と認識した時よりも過去である。また、「落とした」という出来事の後に「入っていない」という出来事が成立すると判断可能である。このことを踏まえ、本研究では、モダリティに埋め込まれた出来事表現とモダリティ表現を別々の述部として扱う。

アノテーション（分割）対象とするモダリティ表現は、川添らのガイドライン [20] に従い、益岡ら [21] により「真偽判断のモダリティ」と呼ばれているものとし、「なければならない」等の「価値判断のモダリティ」は対象外とする。具体的には、以下の表現を対象とすることとした。

アノテーション対象とするモダリティ表現

「だろう」「であろう」「よう」「でしょう」「ちがいない」「間違いない」「疑いない」
「疑いようもない」「はずだ」「はず」「そうだ」「では」「ではないか」「のではないか」
「のじゃないか」「のでは」「のではないだろうか」「のではないかしら」「かもしれない」
「のだ」「かも」「かもだ」「かも分からない」「(し) かねない」「あり得る」「あり得ない」
「という」「らしい」「そうだ」「ようだ」「みたいだ」「っぽい」「まい」「わけではない」
「のではない」「ということだ」「はずはない」「はずがない」「わけはない」「わけがない」

なお、「価値判断のモダリティ」をアノテーション対象外とする理由は、真偽判断のモダリティの場合と異なり、価値判断のモダリティに埋め込まれた出来事表現と、モダリティの外側の出来事表現の時間関係の判断が難しいケースが多いためである。以下に例を示す。

- (23) a. テストの点数が 悪かった。
b. もっと 勉強する べきだった。

ここで、「勉強する」は実際には成立していないため、「勉強する」と「悪かった」の成立順序の判断はできないと考えられる。故に、本研究では「真偽判断のモダリティ」のみ対象とすることとした。

3.2.3 活用形・時制ラベルの付与方法

3.2.1 節で定義した「述部」に対して、活用形と時制ラベルをアノテーションする。1つの述部につき1ラベルを付与する。各ラベルの定義は以下を参照。

Term-Past (終止形-過去)

終止形で終わる用言・モダリティのうち、過去形のものに対してこのラベルを付与する。

- (24) 昨日は 寒かった。

Term-NonPast (終止形-非過去)

終止形で終わる用言・モダリティのうち、非過去形のものに対してこのラベルを付与する。

- (25) 人身事故で電車が 遅延している。

Attr-Past (連体形-過去)

連体形で終わる用言・モダリティのうち、過去形のものに対してこのラベルを付与する。

- (26) 活発で、元気だった 教え子

Attr-NonPast (連体形-非過去)

連体形で終わる用言・モダリティのうち、非過去形のものに対してこのラベルを付与する。

- (27) 気持ちよさそうに 眠る 猫

Cont (連用形)

連用形の用言・モダリティに対してこのラベルを付与する。連用形の時制は未指定であるため、時制情報の付与は行わない。

(28) 論文を 投稿し、採択された。

Te (テ形)

テ形の用言・モダリティに対してこのラベルを付与する。テ形の時制は未指定であるため、時制情報の付与は行わない。

(29) 論文を 投稿して、採択された。

Hyp-Past (条件/仮定形-過去)

条件/仮定形の用言・モダリティのうち、過去形のものに対してこのラベルを付与する。

(30) 昨日、雪が 降ったならば、積もっているはずだ。

Hyp-NonPast (条件/仮定形-非過去)

条件/仮定形の用言・モダリティのうち、非過去形のものに対してこのラベルを付与する。

(31) 明日、雨が 降るならば、試合は中止になるだろう。

Imp (命令形)

命令形の用言に対してこのラベルを付与する。命令形の時制は未指定であるため、時制情報の付与は行わない。

(32) 早く 寝なさい。

Tari (タリ形)

タリ形の用言に対してこのラベルを付与する。タリ形の時制は未指定であるため、時制情報の付与は行わない。

(33) 山を 登ったり、海を泳いだりする。

なお、(33)の「泳いだりする」にも「だり」が含まれるが、終止形で終わっているため、この場合は Term (終止形) ラベルを付与する。

Pre (推量形)

推量形のモダリティに対してこのラベルを付与する。推量形の時制は未指定であるため、時制情報の付与は行わない。

(34) 明日は晴れるでしょう。

(35) やっぱりないのだらうか。

ここで(35)は、2つのモダリティ(「のだ」と「う」)が一塊になった例である。

ModU-Past (ウ接続形-過去)

推量の助動詞「ウ」に接続する用言・モダリティのうち、過去形のものに対してこのラベルを付与する。

(36) 昨日の作業は疲れたらう。

ModU-NonPast (ウ接続形-非過去)

推量の助動詞「ウ」に接続する用言・モダリティのうち、非過去形のものに対してこのラベルを付与する。

(37) 明日は晴れるでしょう。

Ellipsis (省略)

用言の語幹が省略されている場合や語幹のみの場合、このラベルを付与する。省略の場合、時制は未指定であるため、時制情報の付与は行わない。

(38) お問い合わせは、0123-456-789まで。

(39) 帰国後に双子を出産。

Unknown (不明・その他)

上記のどのラベルにも該当しない場合、または、上記のどのラベルに該当するかわからない場合にこのラベルを付与する。

3.3 出来事・モダリティのアノテーション

活用形と時制のラベルを付与した述部に対し、出来事の情報として、event、state、modality のいずれかの属性を付与する。

本研究における「述部」は、出来事か、モダリティを表現している。Kamp ら [11] は、出来事は、event か state のいずれかとしている。述部の出来事が event か state のどちらであるかという情報や、述部が modality であるという情報は、時間情報の解釈や本研究における談話関係の判断に影響を及ぼす。以下に例を示す。

- (40) a. 国境の長いトンネルを 抜ける (非過去の event) と、
b. 雪国であった (過去の state)。

ここで、従属節 (40a) の非過去の event 「抜ける」は、主節 (40b) の過去の state 「雪国であった」が成立している時区間内で成立する、と解釈される。これを図示すると図 3.5 のようになる。



図 3.5 例 (40) の出来事の時区間の配置

続いて次の例を見てみる。

- (41) a. 学校の二階から 飛び降りて (未指定の event)、
b. 一週間ほど腰を 抜かした (過去の event)。

この例は、テ形接続であることから、(41a) の event 「飛び降りる」の成立後に (41b) の過去の event 「抜かした」が成立すると解釈される。これを図示すると図 3.6 のようになる。



図 3.6 例 (41) の出来事の時区間の配置

同様に、(42)を考えてみる。

- (42) a. 路面が濡れている (非過去の state)。
- b. 少し前まで雨が降っていた (過去の state)
- c. ようだ (非過去の modality)。

この例では、(42c)の非過去の modality 「ようだ」は、(42a)の state 「濡れている」が非過去であることから、state 「濡れている」が成立している時区間で成立する、と解釈され、(42b)の state 「降っていた」が過去であることから、state 「降っていた」の終了後に(42c)の modality 「ようだ」が成立すると解釈される。「雨が降る」の成立後に「路面が濡れる」が成立する、という常識的知識も踏まえつつ、これを図示すると図 3.7 のようになる。



図 3.7 例 (42) の出来事の時区間の配置

3.3.1 出来事とモダリティの判定方法

上で述べたように、時間関係や（本研究における）談話関係を解釈する際、述部の出来事やモダリティの情報が手掛かりとなるため、本研究ではこれらの情報のアノテーションを行う。本節では、付与方法の詳細について説明する。

3.2.1 節で述べた通り、述部は、用言またはモダリティである接尾語の語幹に、活用語尾、接尾語の語幹、助動詞が重層的に接続して形成されうる。日本語では、元の用言に助動詞や接尾語が後接することで、eventuality/modality が変化する場合がある。図 3.8 に例を示す。図 3.8 は (42a) の述部「濡れている」の eventuality/modality の変化の例である。ここで、event を指示する動詞「濡れる」に、活用語尾「て」、state を指示する動詞性接尾語の語幹「(て) い(る)」、活用語尾「る」が後接することで、「濡れている」の eventuality が state に変化している。なお、図中の「*」は、直前の eventuality または modality を保持することを表す。

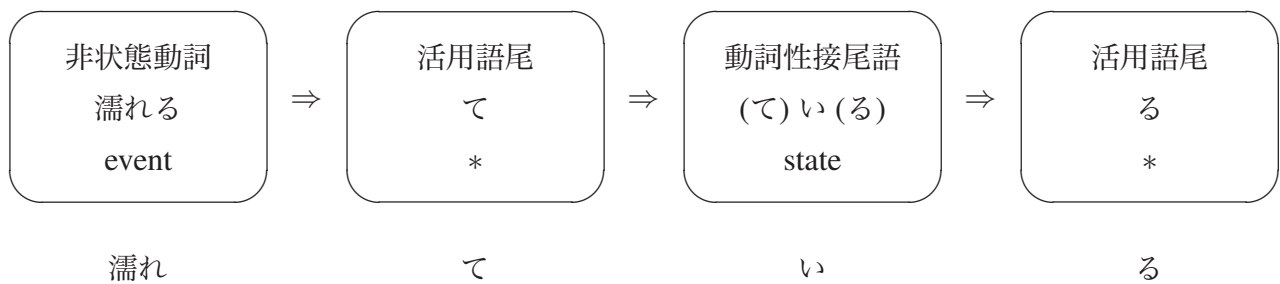


図 3.8 述部「濡れている」の eventuality の変化

本研究では、上述の eventuality/modality の性質を踏まえ、1 つの述部に対し、元の用言の eventuality/modality の属性（以後 init-eventuality と呼ぶ）と、助動詞や接尾語が後接して変化した後の最終的な eventuality/modality の属性を付与する。例えば、(42a) の「濡れている」の場合は、元の用言が非状態動詞「濡れる」であることから init-eventuality として event を付与し、活用語尾や接尾語が後接して state に変化することから、最終的な eventuality/modality として state を付与する。

各述部が event、state、modality のどれを指示するかを判断するため、『日本語文法の形式理論』[19]での品詞分類、および宇津木らの eventuality の分類 [22] に基づき一覧表を作成した（表 3.1）。表中の event、state、modality はそれぞれ、event を指示する語句、state を指示する語句、modality を指示する語句に与えられている。表中の「*」は、図 3.8 と同様に、直前の eventuality または modality を保持することを表す。

品詞分類	述部の分類	eventuality /modality	特徴		例
動詞・語幹	用言・語幹	event	非状態動詞	動きを表す動詞	アル(ク), タベ(ル), ノ(ム)
形容詞・語幹		state	状態動詞	状態を表す動詞	ア(ル), イ(ル), デキ(ル)
					アカ(イ), ウツクシ(イ)
状詞・語幹					キレイ(ダ), シズカ(ダ)
動詞性 接尾語・語幹	モダリティでない 接尾語・語幹	event	形式動詞	動きを表す形式動詞	～(スル), ～ナ(ル)
		state		状態を表す形式動詞	～イ(ル), ～ア(ル)
		event	動詞の連用形 に接続	アスペクトに関係	～ハジメ(ル), ～ダ(ス)
				完遂の意味を表す	～ツク(ス), ～ヌ(ク)
				不完遂の意味を表す	～ワスレ(ル), ～ソコナ(ウ)
				その他	～ア(ウ), ～ナオ(ス)
		state	動詞のテ形 に接続	アスペクトに関係	～(テ)イ(ル), ～(テ)ア(ル)
		*		授受に関係する	～(テ)モラ(ウ), ～(テ)ア(ル)
				その他	～(テ)オ(ク), ～(テ)ミ(ル)
		event	普通名詞に接続		～メ(ク), ～バ(ム)
		*		尊敬	～レ(ル)
		event		受身	～レ(ル), ～(ル)
				使役	～セ(ル)
		state		可能	～(ル), ～レ(ル), ～(ウル)
形容詞/名詞の語幹に接続			～ガ(ル)		
state	否定の意味を持つ		～ナ(イ)		
形容詞性 接尾語・語幹	モダリティである 接尾語・語幹	modality	動詞の連用形に接続	～タ(イ)	
			ダロウ接続形に接続	～ラシ(イ)	
			イ形容詞/状詞語幹に接続	～ッポ(イ)	
			名詞に接続	～クサ(イ)	
			動詞の終止形に接続	～ベ(シ)	
			テ形に接続	～ラシ(イ)	
			その他	～カモシレナ(イ)	
状詞性 接尾語・語幹			～ヨウ(ダ), ～ハズ(ダ)		
助動詞	モダリティでない 助動詞	*	過去の意味を表す	～タ, ～ダ	
	state	断定/丁寧/否定の意味を表す	～(デス), ～ナ(イ), ～ズ		
	モダリティである 助動詞	modality	推量の意味を表す	～ダロウ, ～マイ, ～ン	
活用語尾		*	用言/接尾語の語幹に接続	(タベ)ル, (アカ)イ	

表 3.1 eventuality/modality 分類表

3.3.2 非状態動詞と状態動詞の判別方法

表 3.1 に見られるように、本研究では、動詞は event を指定する動詞（非状態動詞）と state を指定する動詞（状態動詞）に分けられる。本節では、本研究における非状態動詞と状態動詞の判別方法について述べる。

金田一 [23] は、接尾語「ている」との接続可能性と「ている」の表す意味の特徴に基づいて、日本語の動詞を以下のように分類している。

- 状態動詞：状態を表し、「ている」を後接することのない動詞。「ある」「いる」「要する」など。
- 継続動詞：動作・作用を表し、「ている」形式のとき、その動作が進行中であることを表す動詞。「読む」「書く」など。
- 瞬間動詞：動作・作用を表し、「ている」形式のとき、動作・作用が終わってその結果が残存していることを表す動詞。「死ぬ」「消える」など。
- 第四種動詞：常に「ている」が後接した形式で用いられ、状態を帯びることを表す動詞*4。「聳える」「ありふれる」など。

しかし、実テキストにおける動詞を、上述の分類に基づいて区別する作業は人手にとっても機械にとっても自明ではない。まず、状態を表す動詞の数は、上記文献で例として挙げられているものより遙かに多い。また、多くの動詞には用法上の曖昧性があり、ある用法においては瞬間動詞であるものが、他の用法において第四種動詞である、というような例が多く存在する。以下にその一例を示す。

(43) 雪の重みで竹が曲がっている。

(44) 道が曲がっている。

ここで (43) の「曲がっている」は動作・作用の結果を表しており、この「曲がる」は瞬間動詞である。一方 (44) の「曲がっている」は状態を表しており、この用法での「曲がる」は常に「ている」が後接する形で用いられるため第四種動詞である。

そこで本研究では次頁のような判別テストを定義し、それに基づいて動詞の eventuality を判断した。

*4 主文については「ている」は必須となるが、「ありふれた歌詞の曲」等、関係節ではその限りではない。

動詞の eventuality 判別テスト

1. テスト対象の動詞に対して「ている」を後接することができるか。後接できなければ state と判断する。
2. 「<テスト対象の動詞の非過去形>。今は <テスト対象の動詞のテ形> いないけれど。」と述べた際、1 文目と 2 文目が矛盾するかどうか。矛盾すれば state、矛盾しなければ event と判断する。1 で「ている」を後接できた動詞のみ、このテストを行う。

2 つ目のテストは、「非過去形の event は未来に成立し、発話時点では成立しない」という性質と「非過去形の state は発話時点に成立する」という性質を利用している。以下に 2 つ目のテストの実施例を示す。

(45) 部屋を掃除する。今は掃除していないけれど。

(46) 気まずいムードが漂う。#今は漂っていないけれど。

(45) の 1 文目の「掃除する」は未来についての出来事であり、発話時点で成立していると解釈することはできない。したがって、2 文目の「今は掃除していない」と矛盾せず、「掃除する」は event と判断される。一方、(46) の 1 文目の「漂う」は発話時点で成立しているように読めるため、2 文目の「今は漂っていない」と矛盾する。故に、「漂う」は state と判断される。続いて以下の例を見てみる。

(47) これから重たい荷物を持つ。今は持っていないけれど。

(48) ピアニストの素質を持つ。#今は持っていないけれど。

(47) の 1 文目の「持つ」は「手に取る」の意味で用いられており、発話時点では成立しない出来事であると解釈される。そのため、(45) と同様に event と判断される。一方、(48) の 1 文目の「持つ」は「ある性質・状態などを有する」の意味で用いられており、発話時点で成立しているように読める。故に、(46) と同様に state と判断される。このように、対象の述部が event を表す語義と state を表す語義の両方を持つ場合があり、どの語義で用いられるかは文脈によって変わるため、このテストは文章中での語義に合わせて行うこととした。

3.3.3 Ellipsis の述部における出来事とモダリティの判定方法

ここまで、述部に対する、出来事・モダリティの判定方法について述べてきたが、Ellipsis という活用形・時制ラベルが付与された述部では、判定に必要な活用形・時制の情報も省略されてし

まっているため、上記の方法では判断できない。故に、本節では、Ellipsis が付与された述部における出来事・モダリティの情報の判定方法について述べる。本研究では、アノテーション対象の文章の文脈に合わせて、以下のように判断するとした。

- 述部が event を指定するサ変動詞の語幹となり得る形式で、したがって「する」「した」を後接可能な場合は event と判断する。
- 上記以外の場合は state と判断する。

以下に例を示す。

- (49) a. 迷子犬を 発見。
b. これから警察に連れて行きます。
- (50) a. 今日は 夏休み最終日。
b. しかし、宿題は終わっていない…。
- (51) a. お問い合わせは hoge@fuga.co.jp まで。
b. ご返信まで 2-3 日かかる場合があります。

ここで、(49) の「発見」は「発見した」の省略と考えられる。「発見する」は event を指定する動詞なので、event と判断する。一方、(50) の「夏休み最終日」は「夏休み最終日だ」か「夏休み最終日である」の省略で、event を指定する動詞の語幹ではないと考えられるため、state と判断する。(51) は、用言の語幹も省略されてしまっているケースである。このような場合は一律で state と判断することとする。

3.4 述部間の接続先の判定方法

ここでは、本研究における、談話関係を付与する述部対の判断方法について述べる。DGB のように、隣接する談話間に限定しないとすると、アノテーションの作業負担が大きくなり、また、アノテータが異なる接続先を選択してしまった場合、その後の談話構造が大きく異なってしまうという問題がある。この問題は、SDRT のように RFC を採用することで部分的に解決されるが、一方で、接続先は同じでも水平・垂直が異なるラベルを付与してしまう場合があり、問題を拡大することがある。以上を踏まえ、以下のように定義した。

- (a) 従属節（関係節*⁵を除く）の述部の場合、係り受け関係にある述部に接続する。係り元の述部から係り先の述部にエッジを張る（例は図 3.9）。

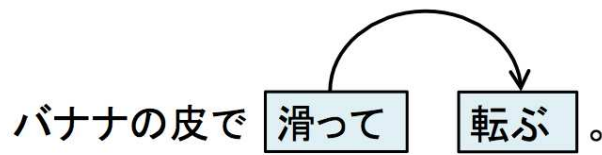


図 3.9 (a) の例

(a') 従属節内の述部動詞が係り受け関係にある場合も (a) に従う。関係節内の述部と関係節外の述部は接続させない (接続させる例は図 3.10、接続させない例は図 3.11)。

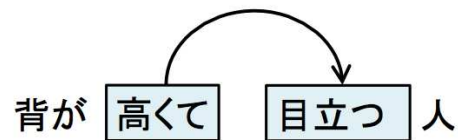


図 3.10 (a') の「関係節内部の述部同士が係り受け関係にある場合」の例



図 3.11 (a') の「関係節内の述部と関係節外の述部」の例

(b) 主節の述部の場合、先行する隣接文の主節の述部に接続する。前の文の述部から後の文の述部にエッジを張る (例は図 3.12)。

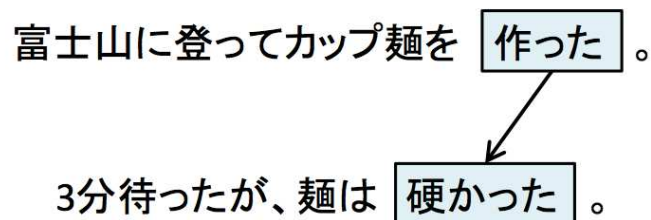


図 3.12 (b) の例

*5 本研究では、「係り先の名詞と内の関係を持つ連体節」を「関係節」と呼ぶ。厳密には関係節という用語は、英語文法等において関係代名詞によって導かれる形容詞節を表すものであるため、関係代名詞は日本語においては陽に存在するわけではない。故に議論の分かれるところではあるが、ここでは便宜上そう呼ぶ。

- (c) 複数文からなる引用節と、それを埋め込む述部が存在する場合はそれらを一塊の談話とみなし、その前（または後）の文や節の述部と接続させる（例は図 3.13）。

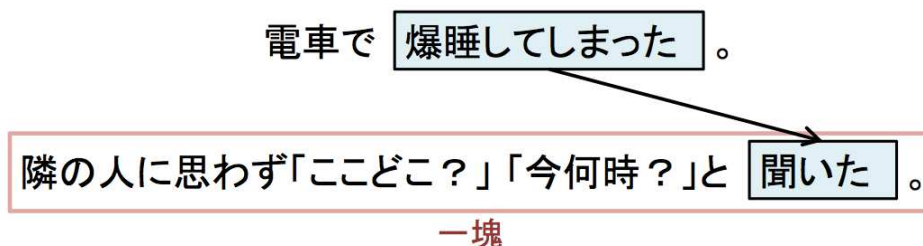


図 3.13 (c) の例

図 3.13 では引用節を埋め込む述部とその前の文の述部間の接続関係しか示していないが、実際は、「一塊の談話」の中にエッジを張れる文や節があればそれらも接続する。上記の (c) の例の場合は、(b) の定義により、図 3.13 の接続関係の他、「どこ?」と「何時?」の述部間、(a) の定義により、「思わず」と「聞いた」の述部間を接続する。

- (d) 少なくとも一方の文がモダリティを伴う場合は、モダリティと前（または後）の文の主節の述部、モダリティに埋め込まれた述部と前（または後）の文の主節の述部をそれぞれ接続させる。エッジの向きは文の順序に合わせる（例は図 3.14、図 3.15）。

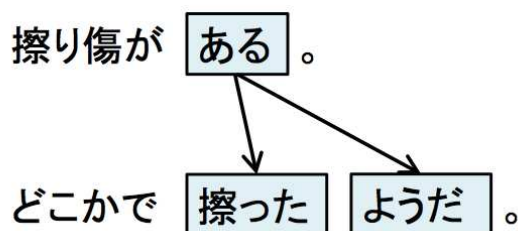


図 3.14 一方の文にのみモダリティがある場合の例

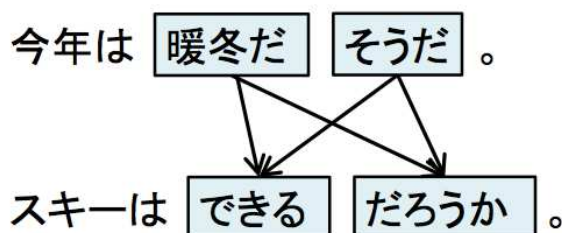


図 3.15 (モダリティがある文同士で接続する例)

図 3.14 は一方の文にのみモダリティがある場合の例であり、図 3.15 はモダリティがある文同士の文間で接続する例である。図 3.15 の場合は、前の文の主節のモダリティと後の文の主節のモダリティ間、前の文の主節のモダリティと後の文の主節のモダリティに埋め込まれた述部との間、前の文の主節のモダリティに埋め込まれた述部と後の文の主節のモダリティとの間、前の文の主節のモダリティに埋め込まれた述部と後の文の主節のモダリティに埋め込まれた述部との間の 4 つにエッジが張られることになる。

3.5 談話関係アノテーション

本研究では、出来事とモダリティの情報、時制の情報、接続表現を手がかりとして、客観的に判断できるよう談話関係を定義する。談話関係は、「命題*6間の時間的な関係 (Temporal Relations)」「明示的に因果を主張する関係 (Causal Relations)」「論理的な関係 (Logical Relations)」のいずれかであるとする。

3.5.1 出来事時、視点時、発話時、参照時

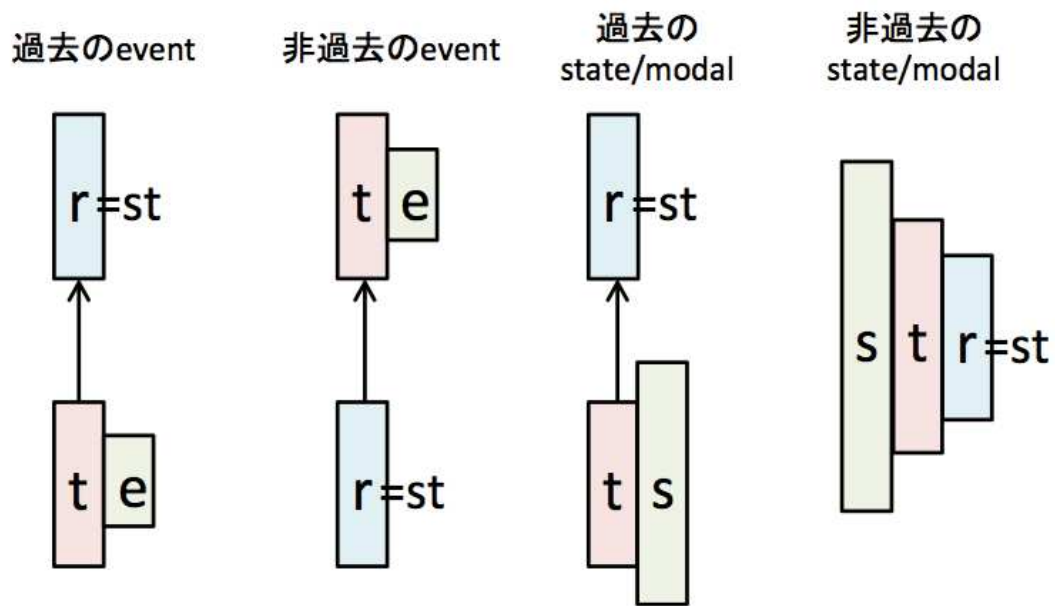
まず本節で「命題の時間的な関係 (Temporal Relations)」を定義する上で必要となる概念について説明する。

単文の発話時 st 、参照時 r 、視点時 t 、出来事時 e 、 s の時間配置は、日本語の時制についての先行研究 [22][25] に従えば、以下のようにになると考えられる。

ここで、出来事時 e 、 s および視点時 t 、発話時 st 、参照時 r の定義は以下の通りである。

- 出来事時 e は、event e が生じてから終わるまでの時区間を表し、出来事時 s は、state/modality s が成立している時区間を表す。
- 視点時 t は「event に対しては少なくともその時区間中に起こったという形で、state/modality については少なくともその時区間中は成立しているという形で、event や state/modality の時間的配置を制約する時区間」を表す。
- 参照時 r は導入された出来事を時間的に配置する際の基点となる時区間であり、単文の場合や複文の主節の場合は $r = st$ で、複文の従属節の場合は様々である。
- 発話時 st は、「文がいつ発話されたか」を表すもので、原則現在 (now) である。

*6 ここでの命題とは、述部を含む節が表す内容を表す。



注1) st: 発話時、r: 参照時、t: 視点時、e, s: 出来事時
 注2) 矢印の向きは、過去 → 未来

図 3.16 単文の時間配置

以下に、例文とその各述部の出来事時、視点時、参照時、発話時の配置を示す。

- (52) a. 太郎が 来る 前に、
 b. 花子が 来た。

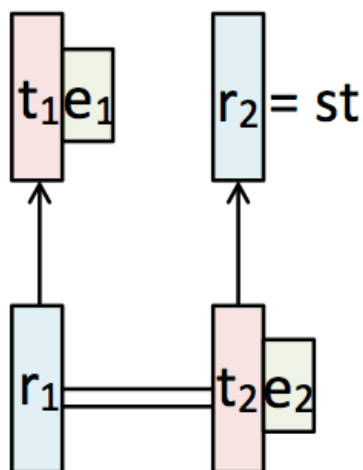


図 3.17 例 (52) の時間配置 1

ここで、 e_1, t_1, r_1 は (52a) の述部「来る」の出来事時、視点時、参照時、 e_2, t_2, r_2 は (52b) の述部「来た」の出来事時、視点時、参照時である。従属節 (52a) の「来る」参照時 r_1 は、主節 (52b) の「来た」の視点時 t_2 になると考えられるため、図 3.17 のような配置となる。なお、図 3.17 は、以下のいずれかの時間配置であることを意味している。図 3.18 の時間配置はそれぞれ、発話時と同時に来た場合、発話時より後に太郎が来ている場合、発話時より前に太郎が来る場合の配置を表している。

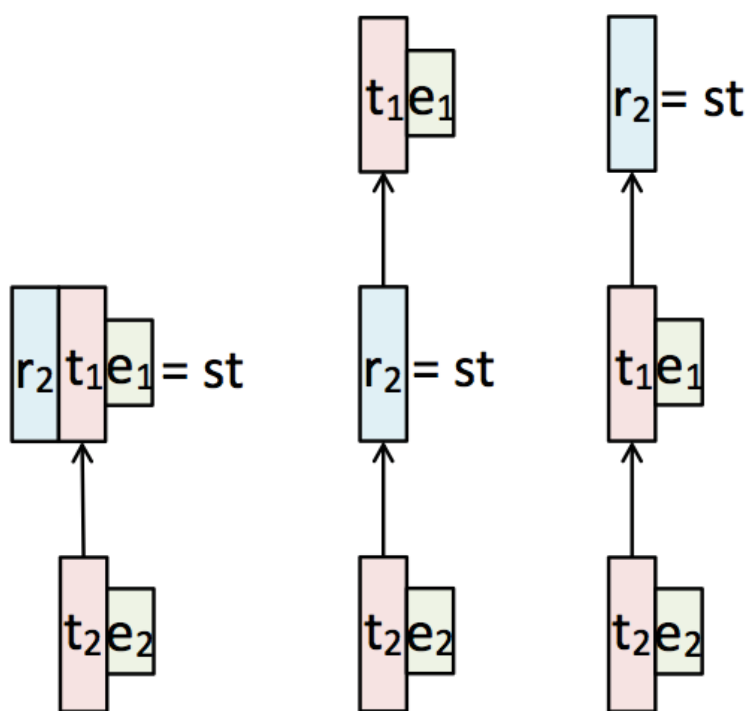


図 3.18 例 (52) の時間配置 2

3.5.2 談話関係ラベルの付与方法

本研究では、1つの述部対に対して、Temporal Relations または Others の中から 1つ、ラベルを必ず付与することとする。述部対が Causal Relations または Logical Relations のいずれかの関係に該当する場合は、そのラベルも追加で付与する。つまり、Causal Relations または Logical Relations のいずれかの関係に該当する述部対の場合、図 3.19 のように Temporal Relations または Others の中から 1つ、Causal Relations または Logical Relations の中から 1つ、ラベルが付与されることになる。

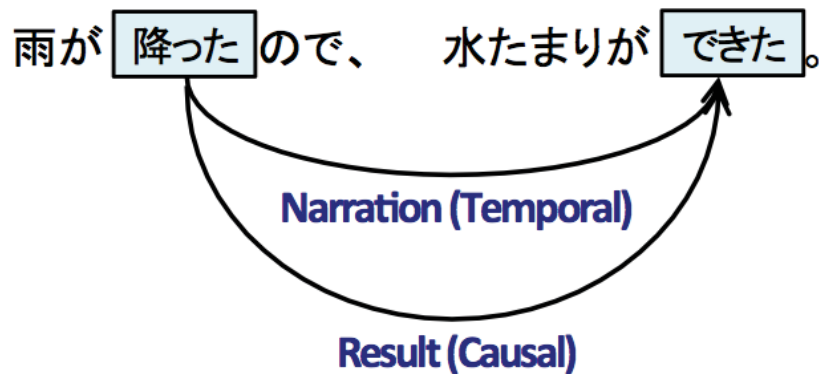


図 3.19 2つの談話関係ラベルが付与される例

以下、各談話関係の定義と付与方法について説明する。

Temporal Relations

命題間の時間的配置に対応する述部間の関係である。本関係では、文章内に明示的に現れる情報（出来事、モダリティ、時制）、および、時間関係アノテーションスキーマである TimeML[26] の TLINK 相当の網羅的な時間関係との依存関係を反映した定義となるよう、2つの述部の視点時の間の関係、出来事時の間の関係、出来事や時制などの属性間の関係を規定している。なお、TimeML とは、時間情報アノテーションの英語のアノテーション基準であり、TLINK とは、時間情報表現と出来事表現の時間的順序関係を関連づけるためのタグセットである。TimeML を元に、多言語へのアノテーションを想定した国際標準 ISO-TimeML[27] が策定されている。ISO-TimeML の日本語適応作業については、浅原ら [28] が行っており、ISO-TimeML を元に 13 種類の時間関係を定義している。図 3.22 に浅原らの時間関係一覧を示す。本研究の談話関係は、多言語に対応できる、汎用的な定義とすることを目指しているものの、日本語の文章を対象としてアノテーションを行う予定であるため、日本語適応を考慮した定義設計となっている。故に、本研究の談話関係は、浅原らが日本語適応作業を施し、定義した時間関係との対応関係が定義されている。

出来事時の間の時間関係との対応関係を適切に反映した談話関係を定義するには、視点時の間の時間関係との対応関係も考慮する必要がある。談話関係の中には、視点時の間の時間関係との対応関係は規定できるが、出来事時の間の時間関係との対応関係は規定できない、という関係も存在するためである。以下に例を示す。

- (53) a. お爺さんは山へ芝刈りに 行った。
 b. お婆さんは川へ洗濯に 行った。

(53)の「山へ芝刈りに行った」と「川へ洗濯に行った」の出来事時の間の時間関係は、前後の文脈次第で分かることもあるが、少なくともこの文のみでは曖昧である。同時に行った可能性もあれば、どちらかが先に行った可能性もある。一方、(53a)の述部の出来事も(53b)の述部の出来事も過去のeventであり、視点時の時区間は過去の(ほぼ)同じ時区間を持つと考えられる。したがって、(53a)の述部の出来事も(53b)の述部の出来事も、図3.20、図3.21のように、視点時の時区間の範囲内で成立しているとは言えると考えられる。

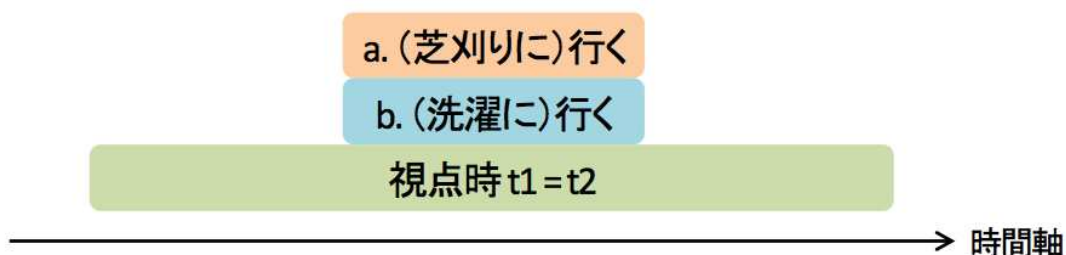


図 3.20 (53) の時間配置の例 1

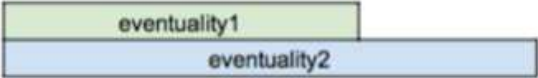
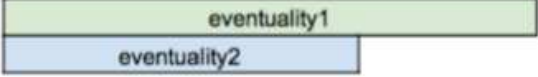
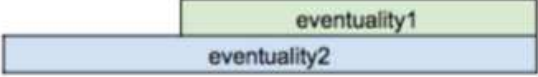
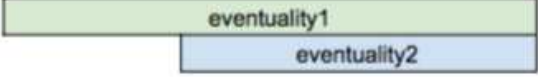
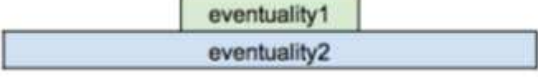
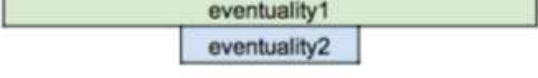
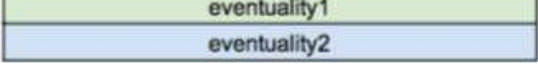
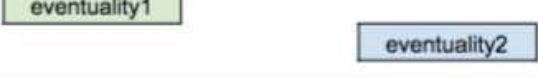


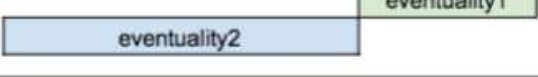
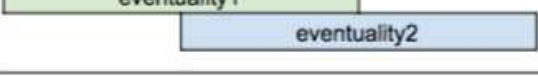
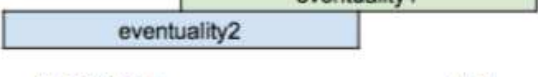


図 3.21 (53) の時間配置の例 2

図3.20、図3.21の視点時 t_1 、 t_2 はそれぞれ1つ目の述部の視点時、2つ目の述部の視点時を表している。このように、(53)のような場合が存在するため、本研究では、出来事時の間の時間関係と談話関係の対応関係だけでなく、視点時の間の時間関係と談話関係の対応関係を考慮した定義となっている。

視点時と出来事時と談話関係の対応関係を表3.2に示す。ここで、 t_1 は1つ目の述部の視点時、 t_2 は2つ目の述部の視点時を表し、 e_1 または s_1 は1つ目の述部の出来事時、 e_2 または s_2 は2つ目の述部の出来事時を表す (e の場合は event、 s の場合は state/modality)。また、時区間 x 、 y に対して、 $x = y$ は「 x と y が同じ時区間であること」、 $x \supset y$ は「 x が y を時間的に包含するこ

と」、 $x < y$ は「 x の開始時は y の開始時より先、かつ、 x の終了時は y の終了時より先、かつ、 x と y の時区間に重なりがないこと」、 $x \circ y$ は「 x の開始時は y の開始時より先、かつ、 x の終了時は y の終了時より先、かつ、 x と y の時区間に重なりがあること」を表している。また、行項目は 1 つ目の述部と 2 つ目の述部の出来事の属性 (1 行目)、視点時の間の時間関係 (2 行目) を表しており、列項目は出来事時の間の時間関係 (浅原らの 13 種類の時間関係) を表している。

概要説明	TLINK タグ	eventuality の時間配置図
開始のみ同時	starts	
	started-by	
終了のみ同時	finishes	
	finished-by	
完全に包含	during/is_included	
	contains/includes	
同時	equal/identity	
前後(重なりなし)	before	
	after	
前後(開始・終了時一致)	meets	
	met-by	
前後(重なりあり)	overlaps	
	overlapped-by	

時間軸: 過去 -----> 未来

図 3.22 浅原ら [28] の時間関係ラベルと出来事の時間配置の対応

		e1-e2			e1-s2					s1-e2					s1-s2	
		t1=t2 / t1ot2 / t2ot1 / t1>t2 / t2>t1	t1<t2	t2<t1	t1=t2	t1<t2 / t1ot2	t2<t1 / t2ot1	t1>t2	t2>t1	t1=t2	t1<t2 / t1ot2	t2<t1 / t2ot1	t1>t2	t2>t1	t1=t2 / t1ot2 / t2ot1 / t1>t2 / t2>t1	t1<t2 / t2<t1
before	e1/s1<e2/s2	Narr	Narr	x	x	Narr	Narr	x	x	Narr	x	x	Narr	x	x	Narr
meets				x	x			x	x		x	x		x		
overlaps	e1/s1o e2/s2		x	x	x	x	x	x	x	x	x	x	x	x	Narr	Narr
contains / includes	e1/s1> e2/s2	Narr / Para	Narr / Para	x	x	x	x	x	Narr / Fore	Narr / Fore	Narr / Fore	Narr / Fore	Narr / Fore	Narr / Fore	Narr / Add	Narr / Add
finished-by				x	x	x	x	x								
starts	e2/s2> e1/s1			x	x	Narr / Back	Narr / Back	x	x	x	x	x	x	x	x	x
equal/identity	e1/s1= e2/s2	Para	x	x	x	x	x	Back	x	x	x	x	x	Fore	Add	Add
started-by	e1/s1> e2/s2	Flash / Para	Flash / Para	x	x	x	x	x	Flash / Fore	Flash / Fore	Flash / Fore	Flash / Fore	Flash / Fore	Flash / Add	Flash / Add	
finishes	e2/s2> e1/s1			x	x	Flash / Back	Flash / Back	Flash / Back								Flash / Back
during / is_included				x	x	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	Flash / Back	
overlapped-by	e2/s2o e1/s1	Flash	Flash	x	x	x	x	Flash	Flash	x	x	x	Flash	Flash	Flash	
met-by	e2/s2< e1/s1			x	Flash	x	x	Flash	Flash	x	x	x	Flash	Flash	Flash	Flash
after				x	Flash	x	x	Flash	Flash	x	x	x	Flash	Flash	Flash	

Para : Parallel, Back : Background, Fore : Foreground, Add : Addition, Narr : Narration, Flash : Flashback

表 3.2 視点時と出来事時と Temporal Relations の対応

以下、Temporal Relations の各関係について述べる。

Narration

述べた順番通りに出来事が成立するような談話関係。以下の条件のうち 1 つ以上を満たしており、かつ、前述の定義に該当する述部間にこのラベルを付与する。

- 1 つ目の述部の出来事時 $e1$ または $s1$ （以降、 $e1/s1$ と記す）と 2 つ目の述部の出来事時 $e2/s2$ について、 $e1/s1 < e2/s2$ 、または、 $e1/s1 \circ e2/s2$ である。すなわち、 $e1/s1$ と $e2/s2$ の間には時間的重なりがないか、 $e1/s1$ が $e2/s2$ よりも先に終了する。
- 少なくともどちらか一方は state である場合で、init-eventuality1 も init-eventuality2 も event であるとき、init-eventuality1 の開始後に init-eventuality2 が開始する。

以下に、Narration が付与される例文と、その時間配置を示す。

- (41) a. 学校の二階から 飛び降りて、
b. 一週間ほど腰を 抜かした。



図 3.23 (41) の時間配置の例

2 つの述部の出来事時に時間的重なりがあり、1 つ目の述部の出来事が明らかに先に開始・終了する（した）と判断できる場合も Narration と判断する。以下に例を示す。

- (54) a. 半年間、就職活動を 行っていた。
b. 今は 2 社の内々定を 保有している。

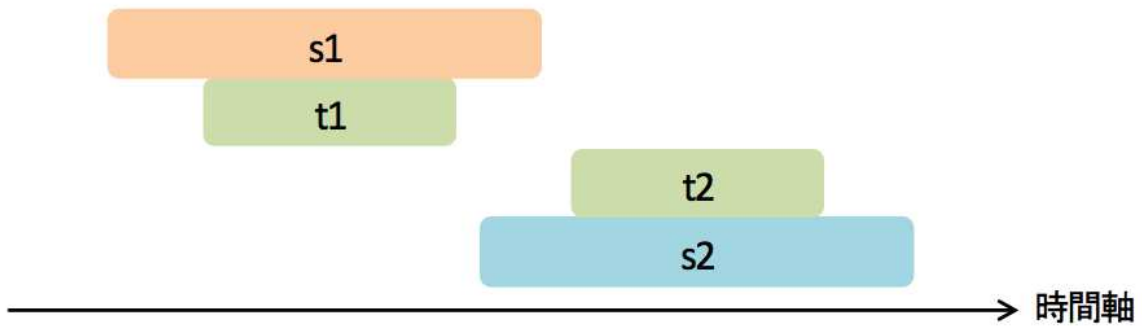


図 3.24 (54) の時間配置

Flashback

Narration の逆。述べた順番とは逆に出来事が成立するような談話関係。以下の条件のうち 1 つ以上を満たしており、かつ、前述の定義に該当する述部間にこのラベルを付与する。

- 1 つ目の述部の出来事時 $e1$ または $s1$ （以降、 $e1/s1$ と記す）と 2 つ目の述部の出来事時 $e2/s2$ について、 $e2/s2 < e1/s1$ 、または、 $e2/s2 \circ e1/s1$ である。すなわち、 $e1/s1$ と $e2/s2$ の間には時間的重なりがないか、 $e2/s2$ が $e1/s1$ よりも先に終了する。
- 少なくともどちらか一方は state である場合で、init-eventuality1 も init-eventuality2 も event であるとき、init-eventuality1 の開始後に init-eventuality2 が開始する。

以下に、Flashback が付与される例文と、その時間配置を示す。(55) は event 同士の関係であるため、図 3.25 のような配置となる。

- (55) a. ご飯を 食べる 前に、
 b. 手を 洗った。

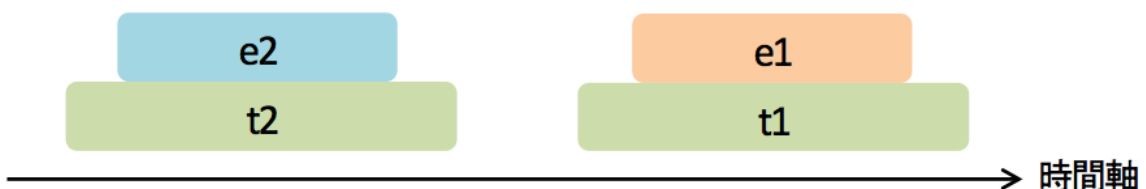


図 3.25 (55) の時間配置

Narration と同様に、2 つの述部の出来事時に時間的重なりがあり、2 つ目の述部の出来事が明らかに先に開始・終了する（した）と判断できる場合も Flashback と判断する。

Background

2 つ目の述部が、1 つ目の述部に対して追加の背景情報を与える談話関係。以下の条件を満たす述部間にこのラベルを付与する。

- 1 つ目の述部の最終的な eventuality/modality は event で、2 つ目の述部の最終的な eventuality/modality は state/modality である。
- 1 つ目の述部の出来事時 $e1$ と 2 つ目の述部の出来事時 $s2$ の間に時間的重なりがある。
- 一方の出来事時が、他方の出来事時を時間的に包含する。つまり、 $s2 \supseteq e1$ または、 $e1 \supseteq s2$ である。

以下に Background が付与される例文と、その時間配置を示す。

- (40) a. 国境の長いトンネルを 抜ける と、
b. 雪国であった。

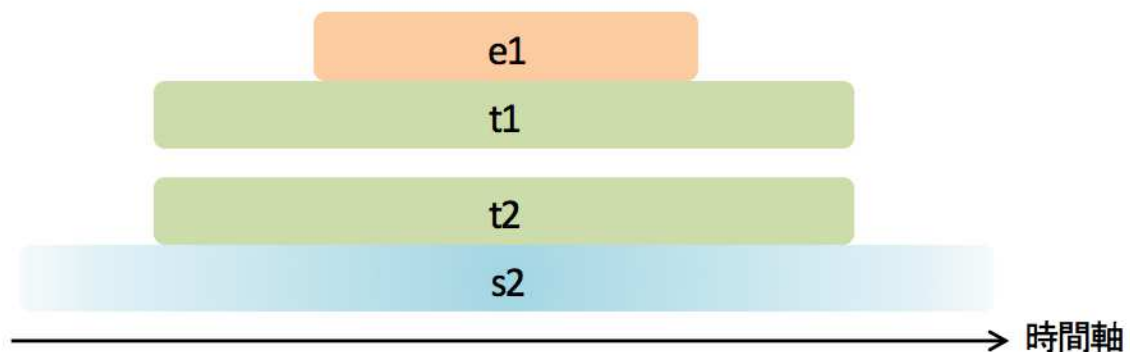


図 3.26 (40) の時間配置

Foreground

Background の逆。1 つ目の述部が、2 つ目の述部の背景情報となるような談話関係。以下の条件を満たす述部間にこのラベルを付与する。

- 1 つ目の述部の最終的な eventuality/modality は state で、2 つ目の述部の最終的な eventuality/modality は event である。

- 1つ目の述部の出来事時 $s1$ と 2つ目の述部の出来事時 $e2$ の間に時間的重なりがある。
- 一方の出来事時が、他方の出来事時を時間的に包含する。つまり、 $s1 \supseteq e2$ または、 $e2 \supseteq s1$ である。

以下に、Foreground が付与される例文と、その時間配置を示す。

- (56) a. 今日はものすごく 寒い。
 b. クローゼットからコートを 出してきた。



図 3.27 (56) の時間配置

Addition

1つ目の述部が表す状態と、2つ目の述部が表す状態の両方が、ある時区間において成立していることを含意する談話関係。以下の条件を満たす述部間にこのラベルを付与する。

- 1つ目の述部と 2つ目の述部どちらも、最終的な eventuality/modality は state または modality である。
- 1つ目の述部の出来事時 $s1$ と 2つ目の述部の出来事時 $s2$ の間に時間的重なりがある。

以下に、Addition が付与される例文と、その時間配置を示す。

- (57) a. 私は器用には 生きられない。
 b. 躓いてばかりいる。

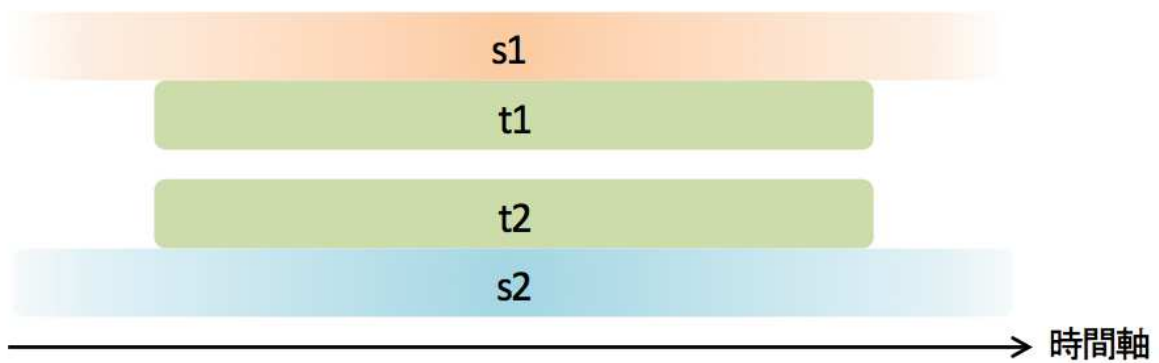


図 3.28 (57) の時間配置

(57) は 2 つの述部の視点時が一致している場合である。以下のように、視点時が一致しないが、2 つの述部の出来事時に時間的重なりがあり、一方の出来事が明らかに先に開始・終了したと判断できない場合も Addition と判断する。

- (58) a. 高校時代、花子はいつでも 元気いっぱい だった。
 b. 今も激務をエネルギーッシュに こなしている。

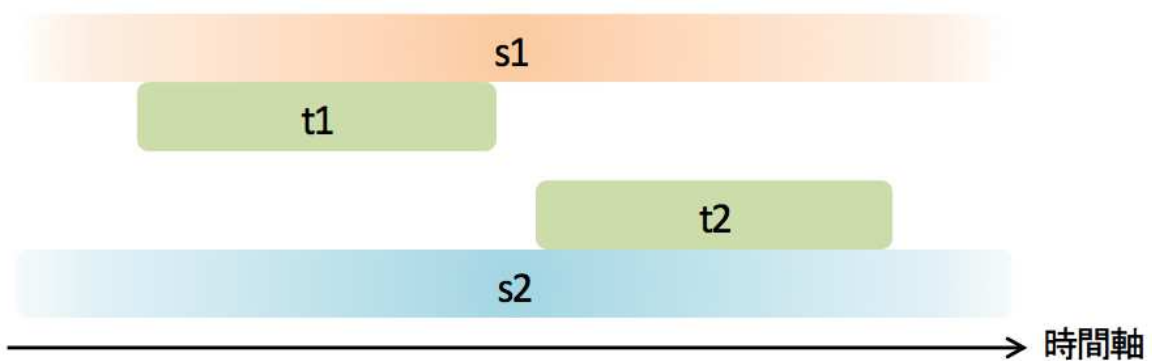


図 3.29 (58) の時間配置

Parallel

Addition と対となる、並列の関係。1 つ目の述部が表す出来事 (event) と、2 つ目の述部が表す出来事 (event) の両方が、ある時区間において成立している場合、または、述部間の視点時の時区間が一致していて、出来事間の成立順序が分からないような場合に該当する。次頁の条件を満たす述部間にこのラベルを付与する。

- 1つ目の述部と2つ目の述部どちらも、最終的な eventuality/modality は event である。
- 1つ目の述部の出来事時 $e1$ と2つ目の述部の出来事時 $e2$ の間に時間的重なりがあり、一方の出来事時が、他方の出来事時を時間的に包含する。または、 $e1$ と $e2$ の開始順序、終了順序が不明確である（2つの述部の視点時の間に時間的重なりがある）。

以下に、Parallel が付与される例を示す。なお、(53) は $e1$ と $e2$ の開始順序、終了順序が不明確である（2つの述部の視点時の間に時間的重なりがある）場合の例であり、図 3.20、図 3.21 などの時間配置を取りうる。

- (53) a. お爺さんは山へ芝刈りに 行き、
 b. お婆さんは川へ洗濯に 行った。

次に、1つ目の述部の出来事時 $e1$ と2つ目の述部の出来事時 $e2$ の間に時間的重なりがあり、一方の出来事時が、他方の出来事時を時間的に包含する場合の例とその時間配置を示す。

- (59) a. フィットネスバイクを 漕ぎ ながら、
 b. テレビを 見る。

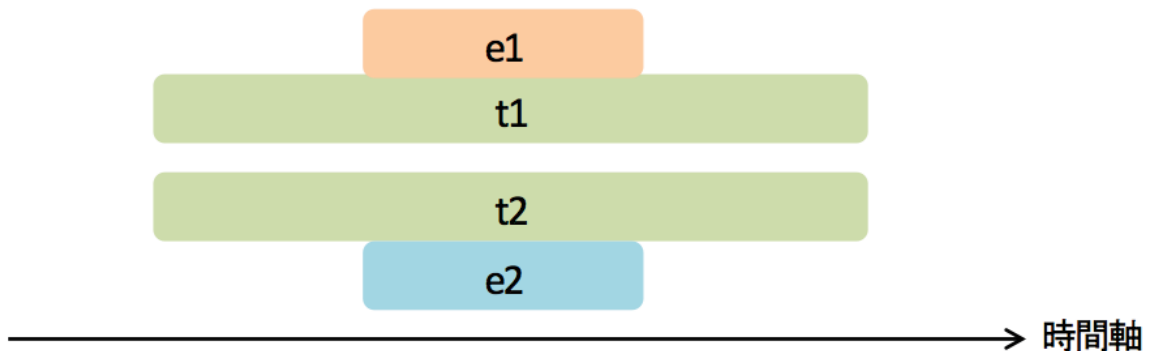


図 3.30 (59) の時間配置

Elaboration

2つ目の述部が、1つ目の述部によって導入される出来事の部分イベントとなっているときに用いる談話関係。Parallel が付与されるものの一部がこの関係に該当する（Parallel の下位ラベルである）。次頁の条件を満たす述部間にこのラベルを付与する。

- 1つ目の述部と2つ目の述部どちらも、最終的な eventuality/modality は event である。
- 2つ目の述部の event が1つ目の述部の event の部分イベントである。
- $e1 \supseteq e2$ である。
- $t1 \supseteq t2$ である。

一方の述部の event が他方の述部の event の部分イベントであるかどうかは、以下のテストにより判断する。

部分イベント判別テスト

以下をすべて満たす場合「述部 A の event が述部 B の event の部分イベントである」と判断する。

- Parallel の条件を満たす。
- 述部 B の動作主 (agents) が述部 A の動作主の部分集合となっていなければならない。
- 述部 B の文に「<述部 A の出来事> (する) 中」「<述部 A の出来事>をする際」といった表現を追加しても文意が通る。
- 述部 A の event (の成立) は述部 B の event (の成立) を含意する。

以下に判別テストの実施例を示す。

- (60) a. 社内展示会を 開催しました。
 b. 自社の新製品の展示紹介や実演が 行われました。
- (61) a. ハンバーグを作ります。
 b. まず玉ねぎをみじん切りにします。

(60) では、(60b) に「開催 ((60a) の出来事) 中」という表現を追加しても文意が通るため、(60b) の述部の出来事は (60a) の述部の出来事である。また、(61) では、(61b) に「ハンバーグを作る際」という表現を追加しても文意が通るため、(61b) の述部の出来事は (61a) の述部の出来事である。

以下に、Elaboration が付与される例文と、時間配置の例を示す。

- (62) a. 今日はカレーを 作ります。
 b. まず野菜を 切ります。

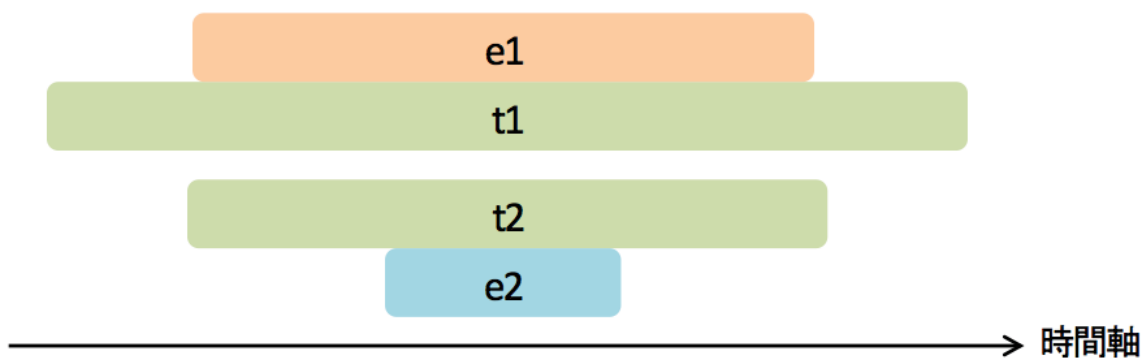


図 3.31 (62) の時間配置

Summary

Elaboration の逆。1 つ目の述部が、2 つ目の述部によって導入される出来事の部分イベントとなっているときに用いる談話関係。Parallel が付与されるものの一部がこの関係に該当する (Parallel の下位ラベルである)。以下の条件を満たす述部間にこのラベルを付与する。

- 1 つ目の述部と 2 つ目の述部どちらも、最終的な eventuality/modality は event である。
- 1 つ目の述部が 2 つ目の述部の部分イベントである。
- $e2 \supseteq e1$ である。
- $t2 \supseteq t1$ である。

以下に、Summary が付与される例文と、時間配置の例を示す。

- (63) a. 火の通った豚肉とキムチを 炒め合わせて、
 b. 豚キムチを 作った。

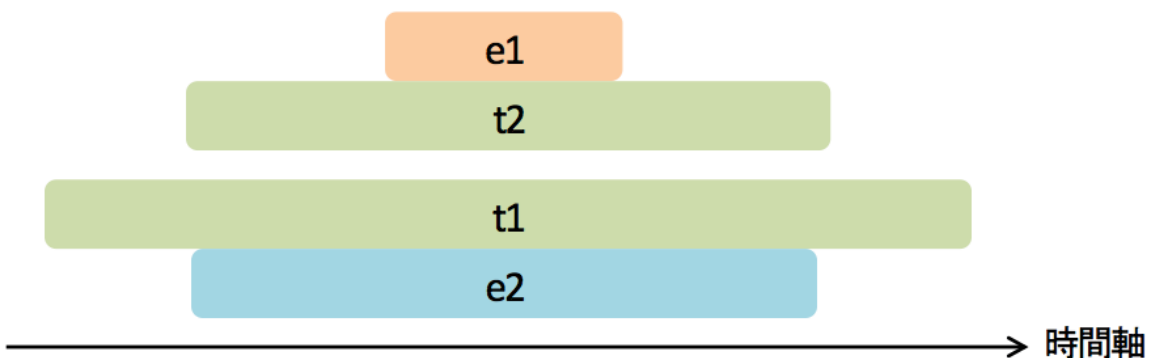


図 3.32 (63) の時間配置

Causal Relations

明示的に因果表現で繋がれた述部間の関係である。因果・理由・目的表現が存在する場合は Causal Relations のいずれかのラベルを必ず付与する。アノテーション対象とする因果・理由・目的表現の例を以下に示す。

アノテーション対象とする因果・理由・目的表現

「～することで」「～することにより」「～することによって」「ので」「のだ」「ため(に)」「から」「～する目的で」「～することを目的として」「～した結果」「～した理由だ」「～した理由である」「～した影響で」

Explanation

2つ目の述部が、1つ目の述部の原因・理由・目的であることを明示的に述べる談話関係。「ので」「から」等の原因・理由を表す表現で繋がれた節の述部間にこのラベルを付与する。以下に例を示す。

- (64) a. 合格点を 取る ため、
b. 地道に 試験勉強 する。

ここで、「試験勉強する」→「取る」の順序で成立すると解釈できるため、「試験勉強する」と「取る」との間に Explanation を付与する。次に以下の例も見てみよう。

- (65) a. 目のクマが 酷い ので、
b. 睡眠時間が 足りなかった のだろう。

(65) では、「睡眠時間が足りなかった」ため「目のクマが酷い」状態になったと判断していると解釈できる。つまり、「足りなかった」→「酷い」の順序で成立している一方、「酷い」→「(足りなかった) のだろう」の順序で成立していると理解できる。故に、「足りなかった」と「酷い」に対して Explanation を付与し、「足りなかった」と「だろう」に対しては Result を付与する。

Explanation は上記の性質から、Temporal Relation が Parallel、Flashback、Background のいずれかの場合に付与される。

Result

1つ目の述部が、2つ目の述部の原因・理由・目的であることを明示的に述べる談話関係。「ので」「から」等の原因・理由を表す表現で繋がれた節の述部間にこのラベルを付与する。

- (66) a. さっき珈琲を 飲んだ ので、
b. 眠くない。

この例では、「飲んだ」から「眠くない」状態になったと判断していると解釈できる。つまり、Explanation と異なり、出来事の成立順序と判断の順序が一致する場合はこのラベルを付与する。

Result は上記の性質から、Temporal Relation が Parallel、Narration、Foreground のいずれかの場合に付与される。

Logical Relations

「または」「ならば」などの論理的な表現で繋がれた述部間の関係である。

Alternation

「A または B」のように、1つ目の述部と2つ目の述部が or の関係となっているもの。「または」「あるいは」等の選択的表現で繋がれた節の述部間にこのラベルを付与する。

- (67) a. 建売住宅を 買う か、
b. 賃貸物件を 借りる か。

選択的表現が存在する場合は必ず付与する。アノテーション対象とする選択的表現の例を以下に示す。

アノテーション対象とする選択的表現

「または」「あるいは」「それとも」「もしくは」「か」「ないしは」

Condition

「もし A ならば B」のように、1つ目の述部が2つ目の述部の条件となっているような談話関係。「ならば」「たら」「すると」等の条件や仮定を表す表現で繋がれた節の述部間にこのラベルを付与する。

- (68) a. 明日雨が 降るならば、
b. 今日のうちに 洗濯してしまおう。

条件・假定表現が存在する場合は必ず付与する。アノテーション対象とする条件・假定表現の例を以下に示す。

アノテーション対象とする条件・假定表現
「ならば」「なら」「ば」「たら」「すると」「しても」

Others

談話関係ではないがアノテーション作業をする上で便宜上必要な関係。

Content

動作とその内容の修飾関係を意味する。具体的には、以下の場合に付与する。

「～と」を用いた引用節の述部と、引用節の係り先の述部間：

- (69) a. 即答を迫るのは 酷だ と
b. 思う。

2つ目の節の述部が、1つ目の節の中の述部以外の要素（名詞など）を詳述している場合で、1つ目の節の述部と2つ目の節の述部の時間関係が判断できない場合：

- (70) a. カレーの作り方について 説明する。
b. まず野菜を 切る。

ここで、(70b)の「切る」は(70a)の「カレーの作り方」を詳述しており、また、(70a)「説明する」との時間関係が判断できないため、Contentを付与する。

「AすることをBする」「AするのをBする」のA、B間：

- (71) a. 全国大会で 優勝する ことを
b. 目指している。

「～するようになる」の「～する」と「なる」の場合：

- (72) a. 規則正しく 生活する ように
- b. なった。

No-Rel

1つ目の述部を含む文や節と、2つ目の述部を含む文や節に繋がりや共通するトピックがなく、談話関係がないと判断される場合や、少なくとも一方の述部が、否定に埋め込まれ、実際は成立していない出来事であるような場合などに付与する。以下に例を示す。

- (73) a. 安易に 認める のではなく、
- b. 慎重に 検討する。

ここで、(73a)の「認める」は否定に埋め込まれており、実際には成立していない。このような場合はNo-Relを付与する。一方、否定形の用言の場合は否定されている状態が成立している時区間が分かるため、それを元に談話関係を判断する。以下に例を示す。

- (74) a. 授業の内容を 理解せず に、
- b. 試験を 受けた。

この場合は「理解しない」状態が成立している間に「(試験を)受けた」と解釈できるので、Foregroundを付与する。

Unknown

上記のどのラベルにも該当しない場合、または、上記のどのラベルに該当するかわからない場合にこのラベルを付与する。

3.5.3 談話関係判断チャート

本研究では、文章内に明示的に現れる情報(出来事、モダリティ、時制)と時間関係、談話関係の決定手順を定義している。次頁にその判断チャートを示す。開始地点の質問から回答を選んで進めていくと、談話関係ラベルが決まるようになっており、各質問・回答は談話関係ラベルの定義と対応するものとなっている。この判断チャートは、談話関係の判断に迷った際の指針となるだけでなく、エラー分析時に役立つと考えられる。

(Q1) 出来事間に時間的重なりがあるか？

(A1-1) はい or 分からない ⇒ (Q2)

(A1-2) いいえ ⇒ (Q5)

(Q2) 出来事の成立（開始）順序は？

(A2-1) e1/s1 → e2/s2 の順で成立（開始）する ⇒ (Q3)

(A2-2) init-eventuality がいずれも event で、init-eventuality1 → init-eventuality2 の順で成立（開始）する ⇒ (Q3)

(A2-3) e2/s2 → e1/s1 の順で成立（開始）する ⇒ (Q4)

(A2-4) init-eventuality がいずれも event で、init-eventuality2 → init-eventuality1 の順で成立（開始）する ⇒ (Q4)

(A2-5) 分からない ⇒ (Q6)

(Q3) e1/s1 が e2/s2 より（明らかに）先に終了するか？

(A3-1) はい ⇒ Narration

(A3-2) いいえ or 分からない ⇒ (Q6)

(Q4) e2/s2 が e1/s1 より（明らかに）先に終了するか？

(A4-1) はい ⇒ Flashback

(A4-2) いいえ or 分からない ⇒ (Q6)

(Q5) 出来事の成立（開始）順序は？

(A5-1) eventuality1 → eventuality2 の順で成立（開始）する ⇒ Narration

(A5-2) eventuality2 → eventuality1 の順で成立（開始）する ⇒ Flashback

(A5-3) 分からない、かつ、Arg1 : event, Arg2 : event ⇒ (Q7)

(A5-4) 分からない、かつ、Arg1 : event, Arg2 : event ではない ⇒ (Q11)

(Q6) Arg1 と Arg2 の eventuality は？

(A6-1) Arg1 : event, Arg2 : event ⇒ (Q7)

(A6-2) Arg1 : event, Arg2 : state/modality ⇒ (Q8)

(A6-3) Arg1 : state/modality, Arg2 : event ⇒ (Q9)

(A6-4) Arg1 : state/modality, Arg2 : state/modality ⇒ (Q10)

(Q7) e1 と e2 の視点時に重なりがあるか？

(A7-1) はい or ないとは言えない ⇒ Parallel

(A7-2) いいえ ⇒ (Q11)

(Q8) s2 が、e1 を時間的に包含するか？

(A8-1) はい or しないとは言えない ⇒ Background

(A8-2) いいえ ⇒ (Q11)

(Q9) s1 が、e2 を時間的に包含するか？

(A9-1) はい or しないとは言えない ⇒ Foreground

(A9-2) いいえ ⇒ (Q11)

(Q10) e1/s1 と e2/s2 は同一の時間軸上にある（何らかの時間関係があると思われる）か？

(A10-1) ある or ないとは言えない ⇒ Addition

(A10-2) いいえ ⇒ No-Rel

(Q11) e1/s1 と e2/s2 は同一の時間軸上にある（何らかの時間関係があると思われる）か？

(A11-1) ある（と思うが、時間配置は分からない） or 分からない ⇒ Unknown

(A11-2) いいえ ⇒ No-Rel

3.6 提案手法の有効性に関する定性的評価

本節では、関連研究 [1, 2, 3, 4] で判断が揺れる例に対して、提案手法に従ってアノテーションした結果を分析し、以下の2点について評価する。

- 談話関係アノテーションの定義から基準のない判断を減らすことができたか。
- 談話関係ラベルから得られる情報は増えたか。

具体的には、2.4 節の例文 (13)(14) における、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果を比較分析することで評価を行う。例文 (13)(14) と、各手法によりそれらの例文にアノテーションを行った結果を次頁に示す。

- (13) a. 眠気が消えて、
 b. やる気が出てきた。

手法名	アノテーション結果	ラベルから得られる情報
RST[1]	cause-result と sequence で揺れると考えられる。(13a) と (13b) の間に因果関係があると考えられるなら cause-result となる。因果関係がない、かつ (13a) の後に (13b) が発生すると考えるならば sequence となる。なお、因果関係の有無を判断するための基準はない。	cause-result : (13a) と (13b) の間に因果関係がある。このラベルからは、どちらが原因、どちらが結果なのかまでは分からない。 sequence : (13a) と (13b) の間に因果関係がない。(13a) の発生後、(13b) が発生する。
DGB[2]	Explanation と Temporal Sequence Relation で揺れると考えられる。(13a) と (13b) の間に因果関係があると考えられるなら、Explanation となる。因果関係がなく、(13a) の後に (13b) が起こると考えるなら Temporal Sequence Relation である。なお、因果関係の有無に関する判断基準はない。	Explanation : (13a) と (13b) の間に因果関係がある。このラベルからは、どちらが原因、どちらが結果なのかまでは分からない。 Temporal Sequence Relation : (13a) と (13b) の間に因果関係がない。(13a) の発生後に (13b) が発生するか、(13b) の発生後に (13a) が発生するかのいずれかである。
PDTB[3]	(13a) と (13b) の間に「したがって」などの因果表現を挿入できると考えるなら、contingency.cause.reason となる。 (13a) と (13b) の間に「その後」などの時間の遷移を示す表現を挿入できると考えるなら、temporal.asynchronous.precedence。 temporal 系のラベルと contingency 系のラベルは同時に付与できる定義になっているため、両方該当すると思われる場合はどちらも付与できる。	contingency.cause.reason : (13a) と (13b) の間に「したがって」などの因果表現を挿入できる。 temporal.asynchronous.precedence : (13a) と (13b) の間に「その後」などの時間の遷移を示す表現を挿入できる。
SDRT[4]	Narration と Result で揺れると考えられる。詳細は 2.4 節を参照。	Narration : (13a) の発生後、(13b) が発生する。(13a) と (13b) の間に時間的重なりはない。 Result : (13a) と (13b) の間に因果関係がある。(13a) が原因、(13b) が結果である。
提案手法	「消える (event)」と「出てくる (event)」の間に時間的重なりがなく、「消える」の開始・終了時は、「出てくる」の開始・終了時より先であると考えると Narration となる。文内に因果表現はないので、Causal Relations は付与しない。	Narration : (13a) 「消える」の開始・終了時は、(13b) 「出てくる」の開始・終了時より先である。このほか、出来事情報や活用形・時制ラベルから「消える」「出てくる」は event であるという情報や、(13b) の出来事表現が過去形であるという情報も得られる。

表 3.3 (13) における、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果

- (14) a. ストープが点いていた。
 b. いつの間にか、猫が焦げていた。

手法名	アノテーション結果	ラベルから得られる情報
RST[1]	cause-result と sequence、temporal-same-time で揺れると考えられる。(14a) と (14b) の間に因果関係があると考えるなら cause-result となる。因果関係がなく、(14a) の後に (14b) が起こると考えるなら、sequence となる。因果関係がなく、必ずしも (14a) の後に (14b) が起こるとは限らないと判断し、「点いている」と「焦げている」の2つの状況が時間的に重なっていると考えるなら temporal-same-time となる。なお、因果関係の有無を判断するための基準はない。	cause-result : (14a) と (14b) の間に因果関係がある。このラベルからは、どちらが原因、どちらが結果なのかまでは分からない。 sequence : (14a) と (14b) の間に因果関係がない。(14a) の発生後、(14b) が発生する。 temporal-same-time : (14a) と (14b) の間に因果関係がない。(14a) と (14b) は時間的に重なっている。このラベルからは、どちらが先に発生したかまでは分からない。
DGB[2]	Explanation と Temporal Sequence Relation で揺れると考えられる。(14a) と (14b) の間に因果関係があると考えるなら、Explanation となる。因果関係がなく、(14a) の後に (14b) が起こると考えるなら Temporal Sequence Relation である。なお、因果関係の有無に関する判断基準はない。	Explanation : (14a) と (14b) の間に因果関係がある。このラベルからは、どちらが原因、どちらが結果なのかまでは分からない。 Temporal Sequence Relation : (14a) と (14b) の間に因果関係がない。(14a) の発生後に (14b) が発生するか、(14b) の発生後に (14a) が発生するかのいずれかである。
PDTB[3]	(14a) と (14b) の間に「したがって」などの因果表現を挿入できると考えるなら、contingency.cause.reason となる。 (14a) と (14b) の間に「その後」などの時間の遷移を示す表現を挿入できると考えるなら、temporal.asynchronous.precedence。 (14a) と (14b) の間に「その間」などの時間の重なりを示す表現を挿入できるならば、temporal.synchronous。 temporal 系のラベルと contingency 系のラベルは同時に付与できる定義になっているため、両方該当すると思われる場合はどちらも付与できる。	contingency.cause.reason : (14a) と (14b) の間に「したがって」などの因果表現を挿入できる。 temporal.asynchronous.precedence : (14a) と (14b) の間に「その後」などの時間の遷移を示す表現を挿入できる。 temporal.synchronous : (14a) と (14b) の間に「その間」などの時間の重なりを示す表現を挿入できる。
SDRT[4]	Narration と Background、Result で揺れると考えられる。Narration と Background で揺れる理由の詳細は 2.4 節を参照。因果関係があると考えるなら、Result となるが、因果関係の有無に関する判断基準はない。	Narration : (14a) の発生後、(14b) が発生する。(14a) と (14b) の間に時間的重なりはない。 Background : (14a) と (14b) の間に時間的重なりがある。 Result : (14a) と (14b) の間に因果関係がある。(14a) が原因、(14b) が結果である。
提案手法	「点いている (state)」と「焦げている (state)」の間に時間的重なりがあり、「点く」の終了時が「焦げる」の終了時より先であるかどうかは分からないと考えると Addition となる。文内に因果表現はないので、Causal Relations は付与しない。	Addition : (14a) 「点いている」は (14b) 「焦げている」と時間的重なりがある。出来事間の開始時・終了時の順序関係は不明である。このほか、出来事情報や活用形・時制ラベルから「点いている」「焦げている」は state であるという情報や、2つの出来事表現が過去形であるという情報も得られる。

表 3.4 (14) における、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果

表 3.3、表 3.4 の結果について議論する前に、まず「ラベルから得られる情報」について考える。ラベル判定が正解した場合、そこから得られる情報というのは、そのラベルがどのような情報量(判断基準)を持っているかということに依存する。例えば RST において、cause-result ラベルが付与されていたら、そこから得られる情報は「因果関係がある」ということである。しかし一方で、RST において、因果関係があるかどうかという判断はアノテータが「因果関係があると思ったかどうか」ということでしかない。異なるアノテータが仮に、共に因果関係があると判断した場合でも、果たして同じ関係についての判断なのかすら保証されていないのである。

PDTB はある意味それを乗り越えようとした。そのために接続詞挿入テストを考案した。これは確かにそのラベルで表される関係が存在するかどうかとは違う基準を与えている。なおかつ、上記で問題となっているような判断の対象に対する不明瞭さはない。ところが、このやり方は本質的な解決になっておらず、A という接続表現が挿入可能なら X という関係であると言えるのかについての分析がない。ということは、PDTB で、A というラベルが付与されていることから得られる情報は、A というラベルが挿入可能であるという情報でしかない。これは談話関係というものの元々のモチベーションから考えると満足できるものではない。

提案手法のアノテーションでは、アノテータの判断には明確なテストが存在しており、これはラベルの意図する関係が存在するかどうかといった同語反復的な基準ではない。したがって、異なるアノテータは少なくとも同じ関係についての判断を行っていることが保証されている。これはその判断が異なる場合においてもそうである。例えば(13)において、「消える」の開始・終了時が「出てくる」の開始・終了時より先であるかどうかという判断は、アノテータの主観によって異なる可能性はある。しかし、このときアノテータは2つの出来事の時間関係を比較するということは理解しており、その基準なしに Narration という名前で表される関係が存在するかどうかと尋ねられているわけではない。

上記の議論を踏まえた上で、(13)(14)のアノテーション結果(表 3.3、表 3.4)を見てみる。まず(13)へのアノテーションは、RST、DGB、SDRT では判断が揺れると推察される。これは、RST、DGB の判断基準、および、SDRT における因果関係の有無の判断基準が同語反復的で、同じ関係についての判断であると保証されていないことに起因していると考えられる。PDTB では「したがって」などの接続表現の挿入可能性から判断するため、判断がある程度一致するものの、前述の通り、接続表現の挿入可能性と時間関係についての分析がない。したがって、アノテーション結果から時間関係についての情報を得ることはできないと考えられる。続いて、(14)の結果を見ると、提案手法以外は判断が揺れると推測されている。これには、前述の問題点だけでなく、関連研究 [1, 2, 3, 4] における時間に関わる談話関係の定義の不備も一因となっていると考えられ

る。RST や SDRT では、「時間的重なりがあるか否か」という判断基準と「どちらの出来事が先に発生するか」という判断基準がそれぞれ独立に定義されている。そのため、(14)の結果のように、時間的重なりがある場合に付与するラベルと、出来事の発生順序が分かる場合に付与するラベルの両方の付与基準を満たしてしまうことがありえる。しかしながら、複数のラベルを同時に付与することは許されていないのである。

提案手法では、因果表現の有無に基づいて因果関係に関わる談話関係 (Causal Relations) を付与する等の定義により、RST、DGB などが抱える同語反復問題を解決している。また、出来事の時間的重なりと出来事の開始時・終了時の順序との依存関係を考慮しながら判断を行うようになっており、関連研究 [1, 2, 3, 4] では揺れていた判断についても、時間に関わる判断が定めれば、談話関係が一意に定まるように定義されている。故に、ここまでの議論とアノテーション結果から、今回の分析の評価は以下のようにになると考えられる。

- 関連研究 [1, 2, 3, 4] で生じていた判断の揺れや、保証されてない定義を解消していることから、提案手法は、関連研究 [1, 2, 3, 4] と比べて「談話関係アノテーションの定義から基準のない判断を減らす」ことができたと考えられる。
- 提案手法では、出来事の時間的重なりと出来事の開始時・終了時の順序との依存関係を踏まえた判断基準を定義していることや、基準のない判断を減らすことができたと考えられることから、関連研究 [1, 2, 3, 4] よりも、付与されたラベルから解釈できる時間的情報は増えたと考えられる。

第 4 章

談話関係アノテーションの実験結果と分析

3 章で提案した手法によりアノテーション実験を行った。本章では、その実験設定を述べ (4.1 節)、実験結果について議論する (4.2 節、4.3 節)。

4.1 実験設定

現代日本語書き言葉均衡コーパス (BCCWJ) [29] コアデータの一部に対して、まず 3.2 節、3.3 節で定義した述部の判定、活用形・時制ラベルの付与と出来事・モダリティの情報の付与を行ったのち、3.4 節、3.5 節の定義に従い、接続先判定および談話関係ラベルの付与を行った。BCCWJ は、現代日本語の書き言葉の多様性を把握するために国立国語研究所で構築されたコーパスで、書籍、雑誌、新聞、白書、Web、法律などから無作為に抽出した約 1 億語のテキストに形態論情報、文書構造タグが付与されたものである。BCCWJ に付与された形態論情報などのアノテーションは、ほとんど自動付与であるが、全体の約 1/100 の量に相当する約 110 万語については、人手により解析精度を高めており、この部分が「コアデータ」と呼ばれている。この中の 10 文書における 482 個の述部対に対してアノテーションを行った。

述部の判定、活用形・時制ラベルと eventuality/modality の情報の付与は、まず 3.2 節、3.3 節の定義に従って付与を行うルールベースの分類器を構築し、BCCWJ のコアデータへ係り受け・並列構造をアノテーションしたデータ (BCCWJ-DepPara) [30] に対して前述の分類器により自動判定を行ったのち、1 名の作業者がアノテーションツール brat[31] 上で人手による修正を施す、という手順で行った。図 4.1、図 4.2 に brat 上での述部アノテーションの様子を示す。

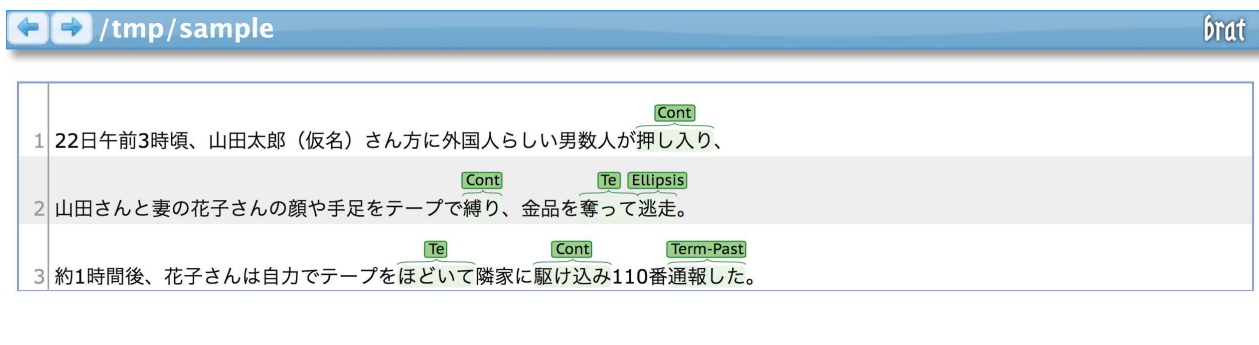


図 4.1 brat における述部アノテーションの様子

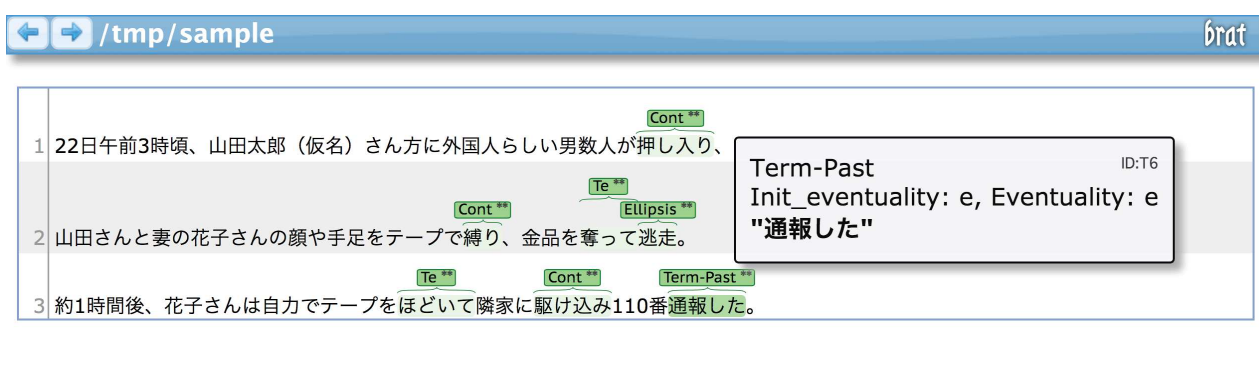


図 4.2 brat 上で述部に付与された eventuality/modality の情報の例

brat では entity ラベルや event ラベルに対して、attribute という属性情報を付与することができる。eventuality/modality の情報は、図 4.2 のように attribute として活用形・時制ラベルに付与することとした。

接続先判定は、述部の判定、活用形・時制ラベルと eventuality の情報を付与済みのデータに対して、brat を用いて 1 名の作業者が行った。談話関係ラベルの付与については、接続先判定まで適用済のデータに対して 2 名の作業者が行った。談話関係ラベルの付与の際、関係ラベルを付与するエッジに対し、どの init-eventuality、最終的な eventuality/modality、時制が対になっているかの情報を図 4.3 のように予め付与しておき、判断しやすくなるようにした。なお、図 4.3 で、赤くハイライトされているものが「どの init-eventuality、最終的な eventuality/modality、時制が対になっているかの情報」である。これらの情報は、ハイフン (-) の左側に 1 目の述部に関する情報、右側に 2 目の述部に関する情報が記述されており、各述部の情報部分には、左から順に、init-eventuality、最終的な eventuality/modality、時制が略記されている。図 4.4 に例を示す。

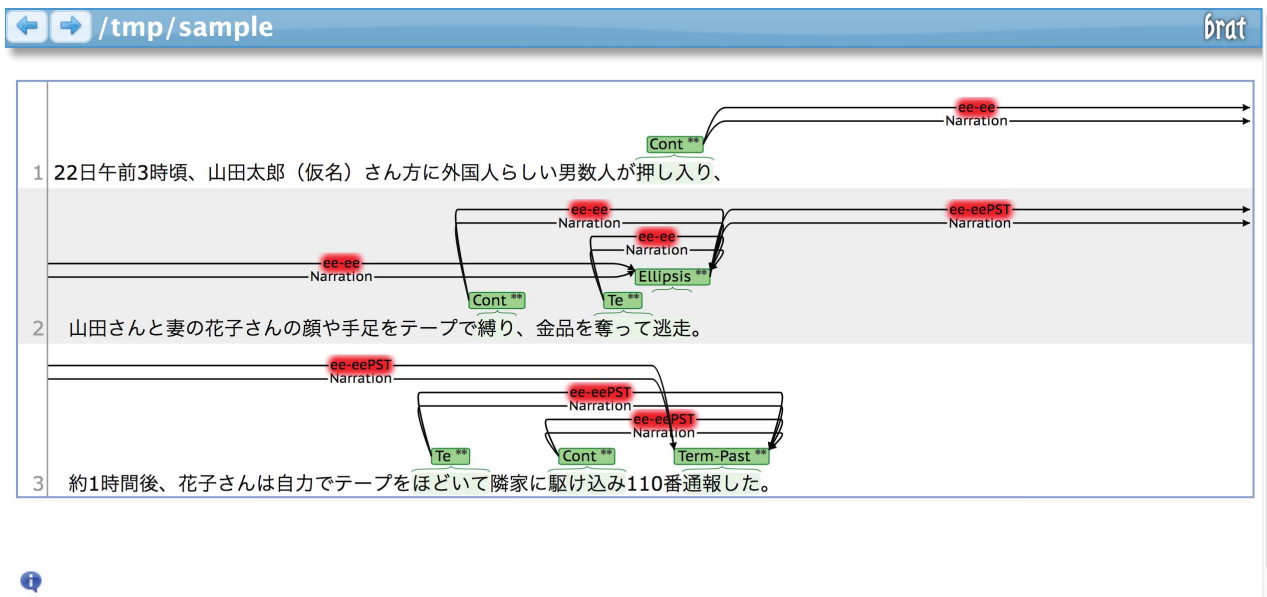


図 4.3 brat における談話関係アノテーションの例

esPST-eNPST

緑文字: **init-eventuality**
 青文字: **eventuality/modality**
 赤文字: **時制**

図 4.4 「どの init-eventuality、最終的な eventuality/modality、時制が対になっているかの情報」の例

ここで、1つ目の述部の init-eventuality は event、最終的な eventuality/modality は state、時制は過去である。e が event、s が state、m が modality、PST が過去、NPST が非過去を表すこととしている。

4.2 談話関係アノテーションの結果

今回扱ったデータにおける一致率は 0.838、カッパ値は 0.817 であった。また、Temporal Relations における一致率は 0.840、カッパ値は 0.809 であり、Causal Relations における一致率は 0.837、カッパ係数は 0.739、Logical Relations における一致率は 0.903、カッパ係数は 0.793 であった。今回のカッパ係数は、カッパ係数の一般判断基準（係数 < 0.4 : poor, 係数 = 0.4~0.6 : fair or moderate, 係数 = 0.6~0.8 : good, 係数 > 0.8 : excellent）からすると、excellent の信頼性を示している。

表 4.1 に本研究における関係ラベルの出現頻度を示す。

		作業 1															
		Add	Back	Flash	Fore	Narr	Para	Elab	Sum	Exp	Res	Alt	Cond	Cont	No	Unk	Total
作 業 者 2	Add	102	0	0	0	1	0	0	0	0	0	0	0	1	2	0	106
	Back	1	52	0	1	6	2	0	0	0	0	0	0	1	0	0	63
	Flash	2	1	9	2	0	3	0	0	0	0	0	0	0	0	0	17
	Fore	0	0	1	67	9	0	0	0	0	0	0	0	0	2	0	79
	Narr	0	3	0	0	78	8	0	0	0	0	0	0	0	1	0	90
	Para	0	0	0	0	14	37	2	0	0	0	0	0	0	0	0	53
	Elab	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	3
	Sum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Exp	0	0	0	0	0	0	0	0	6	0	0	0	0	2	0	8
	Res	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	17
	Alt	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	Cond	0	0	0	0	0	0	0	0	0	0	0	19	0	2	0	21
	Cont	2	1	1	1	1	1	1	0	0	0	0	0	0	51	0	59
	No	0	1	0	1	2	0	0	0	1	3	0	1	0	8	0	17
	Unk	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total	107	58	11	72	112	52	4	0	7	20	1	20	53	17	0	534

表 4.1 本研究における関係ラベルの分布

表 4.1 の関係ラベルは省略表記している。以下に関係ラベル名とその省略表記の対応を示す。

関係ラベル	省略表記	関係ラベル	省略表記
Addition	Add	Explanation	Exp
Background	Back	Result	Res
Flashback	Flash	Alternation	Alt
Foreground	Fore	Condition	Cond
Narration	Narr	Content	Cont
Parallel	Para	No-Rel	No
Elaboration	Elab	Unknown	Unk
Summary	Sum	-	-

表 4.2 関係ラベル名と省略表記の対応表

表 4.1 の結果から、**Addition**、**Content**、**Result**、**Condition**、**Foreground**、**Background** の一致率が高いこと分かる。また、今回、作業 2 名により付けられたラベルが異なっていたペアのうち比較的多かったのは **Narration** と **Parallel** の 22 件、**Narration** と **Foreground** の 9 件、**Narration** と **Background** の 9 件である、ということも確認できる。

続いて、表 4.3 に本研究における関係ラベルと出来事の共起頻度分布を示す。ここで、表 4.3 の各数値は、2 名の作業者の結果を足し合わせたもの（延べ数）を掲載している。また、ハイフン (-) の左側が 1 つ目の述部の属性情報、右側が 2 つ目の述部の属性情報を表しており、出来事情報の各述部には、左から順に、init-eventuality、最終的な eventuality/modality が略記されている。

	ee-ee	ee-es	ee-ss	ee-se	es-ee	es-es	es-ss	es-se	se-ee	se-es	se-ss	se-se	ss-ee	ss-es	ss-ss	ss-se
Add	0	1	0	0	2	19	31	0	0	0	0	0	0	24	134	2
Back	0	27	92	0	0	0	0	0	0	0	2	0	0	0	0	0
Flash	14	1	0	0	0	0	0	0	0	0	0	0	7	3	3	0
Fore	0	0	1	0	17	0	0	0	0	0	0	0	133	0	0	0
Narr	142	15	29	2	3	0	2	0	0	0	0	0	8	0	1	0
Para	98	1	1	1	0	0	0	0	4	0	0	0	0	0	0	0
Elab	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Exp	8	4	1	0	0	0	0	0	0	0	0	0	0	0	2	0
Res	14	0	2	0	0	0	2	0	2	0	0	0	10	0	5	2
Alt	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cond	15	8	13	0	1	0	0	0	0	0	0	0	4	0	0	0
Cont	40	10	27	1	6	3	0	0	2	0	0	0	9	3	11	0
No	6	1	12	0	1	0	1	0	0	0	0	0	5	0	8	0
Unk	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

表 4.3 本研究における関係ラベルと出来事情報の共起頻度分布

表 4.3 の結果から、**Addition** の 5 件 (ee-es : 1 件、es-ee : 2 件、ss-se : 2 件)、**Parallel** の 2 件 (ee-es : 1 件、ee-ss : 1 件)、**Foreground** の 1 件 (ee-ss : 1 件) を除き、各談話関係に課した eventuality/modality の制約を遵守して作業が行われたことが分かる。また、この結果と一致率、カッパ値から、eventuality/modality の制約が談話関係の判断に寄与していると考えられる。

さらに、表 4.4 に本研究における関係ラベルと時制の共起頻度分布を示す。

	UNK-UNK	UNK-PST	UNK-NPST	PST-UNK	NPST-UNK	PST-PST	PST-NPST	NPST-PST	NPST-NPST
Add	18	10	70	2	16	12	7	2	76
Back	6	4	33	4	11	14	7	8	34
Flash	0	4	0	0	3	1	0	17	3
Fore	15	8	37	2	11	14	7	29	28
Narr	16	36	43	8	0	13	26	6	54
Para	16	12	29	4	4	14	2	6	18
Elab	1	0	0	0	0	3	2	1	0
Sum	0	0	0	0	0	0	0	0	0
Exp	0	0	2	0	0	0	0	8	5
Res	1	0	1	2	2	4	0	7	20
Alt	2	0	0	0	0	0	0	0	0
Cond	1	0	8	4	0	0	4	4	20
Cont	2	2	9	7	7	9	15	20	41
No	4	2	6	1	0	0	4	2	15
Unk	0	0	0	0	0	0	0	0	0

表 4.4 本研究における関係ラベルと時制情報の共起頻度分布

表 4.4 の各数値は、2 名の作業者の付与したラベルが一致したものの件数を掲載しており、時制情報の略称表記は、UNK：時制情報なし、PST：過去、NPST：非過去 を表している。表 4.4 の結果から **Narration** が付与された述部対では NPST-PST の属性情報を持つ例が 2 件と少ないこと、一方、**Flashback** が付与された述部対では NPST-PST の属性情報を持つ例が 25 件中 17 件と多く、PST-NPST の属性情報を持つ例が 1 件もないことが読み取れる。また、**Parallel** と **Addition** では、一方が UNK である述部対や、PST-PST、NPST-NPST である述部対が多く (**Addition**：204 件、**Parallel**：155 件)、時制が異なる PST-NPST、NPST-PST である述部対が少ない (**Addition**：9 件、**Parallel**：8 件) ということも分かる。これらから、時制の情報が Temporal Relations の判断に影響を及ぼしていると推察される。

4.3 エラー分析

本節では、談話関係の決定手順のどの部分を明らかにできたかと、どの部分に課題が残るかについて、エラー分析の結果を元に議論する。

まず、Temporal Relations の結果について述べる。Temporal Relations におけるエラーは、(i) 出来事の時区間の解釈が本質的に曖昧で読み手の解釈の違いがアノテーションに反映されたケースが 24 件、(ii) ガイドラインの誤解や単なる付与ミスと思われるケースが 23 件、(iii) 場面や段落の切り替わりの関係で時間関係の判断が難しいケースが 13 件、(iv) モダリティが絡むケースが 9 件、(v) 繰り返しが絡むケースが 4 件、(vi) 一方が **Elaboration** を付与していたケースは 5 件であった。場面の切り替わりや見出し、段落の切り替わりがなく、モダリティや繰り返しもほぼ出現しない 5 文書では Temporal Relations のエラーは見られなかったため、(i)~(vi) に該当しないケースについては、解釈が一意に定まる談話関係やその決定手順を定義できたと考えられる。以下に、(ii) 以外の各ケースの代表例を紹介する。まず (i) の例を示す。

(75) a. 文化庁から国際舞台芸術交流センターへ委任を 行い、

b. 実演デモンストレーション (ショーケース) を 制作しました。 (OW6X_00000^{*7} より)

ここで (75) は、**Parallel** と **Narration** で揺れた例である。「委任」という語には「1. ある物事の処理を他の人にまかせること」「2. 当事者の一方が自分の権利や権限に属する事項の決定や執行を他者にゆだねること、また受任者がこれを受諾することによって成立する契約」という意味がある [32]。したがって、(75a) の「委任」を 1 の意味で「委任を行っている間にデモンストレーション

^{*7} BCCWJ のサンプル ID である。

を制作した」と解釈した場合は **Parallel** となり、2 の意味で「委任契約が受諾された後、デモンストレーションを制作した」と解釈した場合は **Narration** となると判断できる。次に (iii) の例を以下に示す。

- (76) a. 大学に進学する時も、「あなたは論理的に考えるのが得意」と、法学部に行くよう促したのは 池田先生。
b. 大学在学中にデビューし、キャスターとして活躍するその後の進路を思うにつけ、「本当によく見ていてくれた」と 感謝する。(PN1c_00001*⁷ より)

(76) は **Foreground** と **Narration** で揺れた例である。(76a) は大学進学時の場面で、(76b) は、その後暫く経って、過去を回想している場面であり、場面が切り替わっていると考えられる。こうした場面の切り替えを考慮した上で、「池田先生 (だった)」という state の時区間が「感謝する」という event の時区間を包含すると判断したか、前後関係にあると判断したかで談話関係の判断が揺れたと推察される。したがって、場面や段落の切り替わりの際の **Temporal Relations** の判断基準についての定義を見直す必要がある。続いて、(iv) の例を示す。

- (77) a. たつや君はしだいにのめりこんで、四十問も考えた のです。
b. 気がつくとも五回以上も繰り返し 読んでいた のです。(PB43_00001*⁷ より)

(77) は **Addition** と **Flashback** で揺れた例である。(77a) は非過去の modality で、(77b) は過去の state である。(77a) の modality が成立する時区間の解釈、および、(77b) の state の時区間と (77a) の modality の時区間の重なりについての解釈の揺れが結果に反映されていると考えられる。故に、モダリティの時区間やモダリティに関する時間関係の判断基準を明確にする必要がある。次に (v) の繰り返しが絡むケースを以下に示す。

- (78) a. 物語では正解が1つであるわけではありません。
b. 一つの回答があっても、どんどん 「つけたして」 いきます。
c. それが妥当であればすべて一点に します。(PB43_00001*⁷ より)

上記の例は **Parallel** と **Narration** で揺れた例である。(78) では、「(回答を) つけたしていく」という event と「(一点に) する」という event は繰り返されていると考えられるため、繰り返している出来事全体を1つの出来事として捉えると **Parallel** と判断されると思われる。一方、繰り返している出来事1つ1つの時間関係について考えると、「つけたしていく」という event の後に「(一点に) する」という event が成立すると解釈できるため、**Narration** と判断されると推察される。

繰り返しを伴う出来事における関係判断の方法を改めて定義すべきであると考えられる。

さらに (vi) の **Elaboration** に関する結果についても述べる。以下の (79) は、**Narration** と **Elaboration** で揺れた例である。

(79) a. 流行に 乗って

b. かばんを薄くつぶしたり、ピアスを したり。 (PN1c.00001*⁷ より)

ここでまず (79a) と (79b) の動作主は同じである。また、一方の作業者は **Narration** を付与しているが、「流行に乗る中、ピアスをした」と言い換えても文意は通るため、節の部分イベント判別テストの1つ目~3つ目の条件は満たしていると考えられる。しかしながら、「(流行に) 乗る」という出来事に「(ピアスを) する」という出来事を含意するかについては、人によって判断が揺れると考えられる。故に、人によって判断が揺れないもののみテストを通るよう、部分イベントの定義やテストを修正する必要がある。

続いて、**Causal Relations** の結果について議論する。本研究では **Causal Relations** を「明示的な因果表現・理由表現・目的表現で繋がれた述部間の関係」と定義し、代表的な因果・理由・目的表現をいくつかリストアップした上で作業を行った。そのため、「ので」「から」等の代表的な因果・理由・目的表現を伴う文間・節間、および、明らかに因果・理由・目的表現を伴っていない文間・節間における判断はよく一致していた。しかしながら、本研究では、因果・理由・目的表現の列挙を網羅的には行っていなかったため、因果・理由・目的表現自体の判断で揺れが生じ、それがエラーに繋がっていた。具体的には、3.5 節で列挙していなかった「~わけだ」「そこで、~する」などの表現を伴う文間・節間の判断が揺れていた。**Causal Relations** の判断の揺れを減らすには、因果・理由・目的表現の網羅的なリストアップや判断基準を明確に定義する必要があると考えられる。今回の実験では、**Logical Relations** において条件・仮定表現自体の判断での揺れに起因するエラーは確認されなかったが、**Logical Relations** でも同様のエラーが起こることが予想されるため、条件・仮定表現についても網羅的なリストアップが必要であると思われる。

第 5 章

結論

5.1 まとめ

本研究では、談話関係アノテーションの定義から基準のない判断を可能な限り減らすこと、および、時間関係との依存関係を反映した、談話関係とその決定手順を確立することを目指し、次のことを行った。まず、談話関係の判断に関わると考えられ、文章内に明示的に現れる情報（出来事、モダリティ、時制）のアノテーション方法、および、出来事が動作 (event) であるか、状態 (state) であるかを踏まえた談話関係の時間的制約を定義した。その上で、文章内に明示的に現れる情報と網羅的な時間関係との依存関係を反映した談話関係を設計し、それらの決定手順を定義した。その後、関連研究 [1, 2, 3, 4] と本研究のアノテーション結果の定性的評価を行った。その結果、関連研究 [1, 2, 3, 4] と比べ、提案手法はアノテーションの定義から基準のない判断を減らすことができたと考えられること、および、付与されたラベルから解釈できる時間情報が増えたと考えられることが分かった。さらに、定義した分類体系と決定手順に従い、BCCWJ コアデータの 10 文書における 482 個の述部対に対してアノテーション実験を行い、「解釈が一意に定まる談話関係を定義できたか」を検証した。実験の結果、一致率は 0.838、カッパ係数は 0.817 であり、カッパ係数の一般判断基準からすると excellent の信頼性を示していることが分かった。また、談話関係ラベルと出来事、時制の共起頻度分布から、eventuality/modality の制約が談話関係の判断に寄与していること、時制の情報が Temporal Relations の判断に影響を及ぼしていることも確認した。これらの結果は、文章内に明示的に現れる情報に基づいて、客観的に判断可能な談話関係の判断基準を定義できたことを示唆している。

アノテーション実験において不一致だった結果から、談話関係の決定手順のどの部分を明らかにできたか、および、どの部分に課題が残るかについても分析を行った。Temporal Relations におけるエラーは (i) 出来事の時区間の解釈が本質的に曖昧で読み手の解釈の違いがアノテーション

に反映されたケース、(ii) ガイドラインの誤解や単なる付与ミスと思われるケース、(iii) 場面や段落の切り替わりの関係で時間関係の判断が難しいケース、(iv) モダリティが絡むケース、(v) 繰り返しが絡むケース、(vi) 一方が **Elaboration** を付与していたケースのいずれかであることが分かった。このことから、(i)~(vi) に該当しないケースについては、解釈が一意に定まる Temporal Relations やその決定手順を定義できたと考えられる。また、Causal Relations に関しては代表的な因果・理由・目的表現を伴う文間・節間、および、明らかに因果・理由・目的表現を伴っていない文間・節間における判断はよく一致していたものの、3.5 節で列挙されていなかった因果・理由・目的表現における判断で揺れが生じ、それがエラーに繋がっていたことが分かった。

Temporal Relations における (i)~(vi) のケースの判断基準の明確化や、因果・理由・目的表現、条件・仮定表現の網羅的なリストアップなど、まだ多くの課題が残っているが、関連研究が抱える問題の一部については解決することができた。このことから、本研究の試みにより、明示的な情報から客観的に判断可能な談話関係の理論を確立するための土台を作り、計算機に文章・文脈を理解させるという究極目標の実現のための新たな足がかりを築くことができたと考えている。

5.2 今後の課題・展望

今回のエラー分析により判明した課題として、Temporal Relations における (i)~(vi) のケースの判断基準の明確化があるが、いずれも一筋縄ではいかない問題である。以下に、各課題における懸念事項を示す。

- (i) に対処するには、出来事の時区間の解釈の曖昧性を必要十分に吸収できる判断基準や分類体系を定義する必要があるが、それを行うには、アノテーション実験、結果分析と定義の修正を地道に繰り返して決めていかざるを得ないと考えられる。
- (ii) への対処法の1つとして、アノテーションツールで制約違反となるラベルを選択できないようにすることが考えられるが、「制約は違反しているがこの談話関係を付与すべき」といった例外の取扱いを検討した上で導入する必要があると推察される。
- (iii) に対処するには、場面や段落の切り替わりのパターンを列挙し、パターン毎に判断基準を決める必要があると考えられる。というのは、場面が変わり時間も大きく変化したが生きた時間の前後関係は判断できる、という場合もあれば、場面（話題）は変化していないものの出来事間の時間関係は判断できない、という場合もあるためである。
- (iv) に対処するには、各モダリティ表現の分析結果を踏まえてモダリティの時区間やモダリティに関する時間関係の判断基準を決める必要があると思われるが、モダリティ表現自体

の分析が十分に確立しているとは言えないため、まずはモダリティ表現の分析を1つ1つ行う必要があると考えられる。

- (v) への対処は (i) の問題も絡むため難しい。繰り返しを伴う出来事は、「繰り返している出来事全体を1つの出来事として捉えた場合の時間関係」としての側面と、「繰り返している出来事1つ1つの時間関係」としての側面を持つ。しかしながら、一方の談話片対では前者の側面を扱いたい、他方の談話片対では後者の側面を扱いたい、という場合もある。そのため、上述の両側面の共存と切り替えを両立するような定義設計を行う必要がある。
- (vi) への対処のために部分イベントの定義やテストを修正する必要があるが、部分イベントか否かの判断には、作業者の世界知識や常識が絡むため、難しいと考えられる。

理想的な談話関係の理論が確立し、理想的な談話関係認識が出来るようになれば、文章を処理する様々なタスクで役立つ可能性がある。例えば、出来事間の時間関係認識や、出来事の時系列を考慮して返答する質問応答システムや対話システムの研究開発に役立つと考えられる。また、品詞タグ付け、係り受け解析と談話関係認識を同時に行う解析器が作れば、文脈に合わせて曖昧性解消をしながら品詞タグ付けや係り受け解析を行うことができるようになる可能性もある。品詞タグ付け、係り受け解析は、様々な自然言語処理のアプリケーションに関わっているため、このような解析器を構築する意義は大きいと考えられる。このように、談話関係研究の進展は、今後の自然言語処理研究の更なる発展の鍵となると推察される。上で挙げた課題の解決には多大なコストを要すると予想されるが、談話関係の重要性に鑑みれば、これらの課題に取り組む価値は十分にあると考えられる。

謝辞

本研究を進めるにあたり、多くの方々にご協力を頂きました。心より感謝致します。

まず、指導教官であるお茶の水女子大学理学部の戸次大介准教授には、修士1年の夏から今日に至るまで、7年半という長期間にわたって研究活動全般において熱心なご指導、ご助言を頂きました。諸事情により研究が停滞した時も、仕事等の両立ができず悩んだ時も、暖かく見守り、励ましてくださいました。博論提出まで漕ぎ着けられたのは、先生が最後まで見捨てずに面倒を見てくださったからだと思います。本当にありがとうございました。

また、峯島宏次准教授にも、研究活動全般からお悩み相談まで、本当にお世話になりました。お忙しい中、ゼミに付き合ってくださいたり、親身なアドバイス、コメントをたくさん頂きました。戸次研究室、浅井研究室の皆様には日頃から大変お世話になりました。皆様のお陰で、多くのことを学び、楽しく刺激的な大学院生活が過ごせました。心より感謝致します。

ご多忙の中、博論審査をお引き受けくださいました、お茶の水女子大学理学部の小林一郎教授、小口正人教授、浅井健一准教授には、熱心かつ親身なご指導、ご助言を頂きました。また、審査に限らず、授業やゼミ、研究室合宿、学振申請、インターンや就職など、様々な場面でご指導やアドバイス、暖かい励ましのお言葉を頂きました。深く感謝致します。

本研究を進めるにあたり、京都大学の黒橋禎夫教授、河原大輔准教授、東北大学の乾健太郎教授、東京大学の宮尾祐介教授、吉田奈央氏、統計数理研究所の持橋大地准教授、情報通信研究機構の飯田龍氏、科学技術振興機構の建石由佳氏など、ここに書ききれないくらいたくさんの方からアドバイス、鋭いご指摘を頂いたり、CRESTやメタ知識プロジェクトなどに参加し、議論を交わす機会を与えてくださったりしました。ここに感謝の意を表します。

最後に、今日に至るまで支えてくださった家族に心より感謝申し上げます。家族の支えなしでは、家族に背中を押されることがなければ、博士論文を書くことはできなかったと思っています。どうもありがとうございました。

参考文献

- [1] W. C. Mann, and S. Thompson. Rhetorical structure theory: A theory of text organization. Technical report, University of Southern California, Information Sciences Institute, 1987.
- [2] F. Wolf, E. Gibson, A. Fisher, and M. Knight. The Discourse GraphBank: A database of texts annotated with coherence relations. Philadelphia: Linguistic Data Consortium, 2005.
- [3] R. Prasad, A. Joshi, N. Dinesh, A. Lee, E. Miltsakaki, and B. Webber. The Penn discourse treebank as a resource for natural language generation. In *the Corpus Linguistics Workshop on Using Corpora for Natural Language Generation*, pp. 25–32, 2005.
- [4] N. Asher and A. Lascaridas. Logics of Conversation: Studies in Natural Language Processing. *Cambridge University Press*, 2003.
- [5] L. Carlson, D. Marcu, and M. E. Okurowski. RST Discourse Treebank LDC2002T07. Philadelphia: Linguistic Data Consortium, 2002.
- [6] F. Wolf, E. Gibson, A. Fisher and M. Knight. A procedure for collecting a database of texts annotated with coherence relations. *Database Documentation*, 2003.
- [7] M. Marcus, B. Santorini, and M. A. Marcinkiewicz. Building a large annotated corpus of English: The Penn TreeBank. *Computational Linguistics*, Vol. 19, No. 2, pp. 313–330, 1993.
- [8] H. Kūcera, and W. N. Francis. Computational Analysis of Present-Day American English. *Brown University Press*, 1967.
- [9] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The Penn Discourse TreeBank 2.0. In *Proceedings, 6th International Conference on Language Resources and Evaluation*, pp. 2961–2968, 2008.
- [10] R. Prasad, B. Webber, A. Lee, and A. Joshi. The Penn Discourse Treebank 3.0. Philadelphia: Linguistic Data Consortium, 2019.
- [11] H. Kamp and U. Reyle. From Discourse to Logic; Introduction to the Model theoretic Semantics of natural language. *Model theoretic Semantics of natural language*, 1993.

- [12] B. Reese, J. Hunter, N. Asher, P. Denis, and J. Baldridge. Reference manual for the analysis and annotation of rhetorical structure (version 1.0). Technical report, Austin: University of Texas, Departments of Linguistics and Philosophy, 2007.
- [13] D. Kawahara, Y. Machida, T. Shibata, S. Kurohashi, H. Kobayashi, and M. Sassano. Rapid Development of a Corpus with Discourse Annotations using Two-stage Crowdsourcing. In *Proceedings of the 25th international conference on Computational Linguistics*, pp. 269–278, 2014.
- [14] Y. Kishimoto, S. Sawada, Y. Murawaki, D. Kawahara, and S. Kurohashi. Improving Crowdsourcing-Based Annotation of Japanese Discourse Relations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pp. 4044–4048, 2018.
- [15] 岸本裕大, 澤田晋之介, 村脇有吾, 河原大輔, 黒橋禎夫. クラウドソーシングを用いた談話関係アノテーションの改良. 言語処理学会第 23 回年次大会発表論文集, pp. 819–822, 2017.
- [16] K. Murakami, E. Nichols, S. Matsuyoshi, A. Sumida, S. Masuda, K. Inui, and Y. Matsumoto. Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments. In *Proceedings of the 3rd Workshop on Information Credibility on the Web*, pp. 43–50, 2009.
- [17] J. Bos, S. Clark, M. Steedman, J. R. Curran, and J. Hockenmaier. Widecoverage semantic representations from a CCG parser. In *Proceedings of the 20th international conference on Computational Linguistics*, pp. 1240–1246, 2004.
- [18] P. Martínez-Gómez, K. Mineshima, Y. Miyao, and D. Bekki. ccg2lambda: A Compositional Semantics System. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics - System Demonstrations*, pp. 85–90, 2016.
- [19] 戸次大介. 日本語文法の形式理論: 活用体系・統語構造・意味合成. 第 24 巻, くろしお出版, 2010.
- [20] 川添愛, 齊藤学, 片岡喜代子, 崔榮殊, 戸次大介. 言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4. *Technical Report of Department of Information Science, Ochanomizu University, OCHA-IS 10-4*, 2011.
- [21] 益岡隆志. 日本語モダリティ探究. くろしお出版, 2007.
- [22] 宇津木舞香. 依存型理論による日本語のテンスの分析. お茶の水女子大学 大学院人間文化創成科学研究科 修士論文, 2017.
- [23] 金田一春彦. 国語動詞の一分類. 言語研究 15, 1950.
- [24] 寺村秀夫. 日本語のシンタクスと意味 II. くろしお出版, 1984.

- [25] 中村ちどり. 日本語の時間表現. 第 14 巻, くろしお出版, 2001.
- [26] J. Pustejovsky, J. M. Castano, R. Ingria, R. Sauri, J. Robert, G. A. Setzer, G. Katz, and D. R. Radev. TimeML: Robust specification of event and temporal expressions in text. *New directions in question answering*, Vol. 3, pp. 28–34, 2003.
- [27] J. Pustejovsky, K. Lee, H. Bunt, and L. Romary. ISO-TimeML: An International Standard for Semantic Annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, pp. 394–397, 2010.
- [28] M. Asahara, S. Yasuda, H. Konishi, M. Imada, and K. Maekawa. BCCWJ-timebank: Temporal and event information annotation on Japanese text. In *the 27th Pacific Asia Conference of Language Information and Computation*, pp. 206–214, 2013.
- [29] K. Maekawa. Balanced Corpus of Contemporary Written Japanese. In *Proceedings of the 6th Workshop on Asian Language Resources (ALR-8)*, pp. 101–102, 2008.
- [30] 浅原正幸, 松本裕治. 『現代日本語書き言葉均衡コーパス』に対する係り受け・並列構造アノテーション. 言語処理学会第 19 回年次大会発表論文集, pp. 66–69, 2013.
- [31] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii. BRAT: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 102–107, 2012.
- [32] 松村明. 「委任」. 大辞林第 3 版, 三省堂, 2006.

発表文献一覧

査読付き国際会議・ワークショップ

1. Kimi Kaneko, Daisuke Bekki. “Building a Corpus of Temporal-Causal-Discourse Structures Based on SDRT for Extracting Causal Relations,” In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pp.33–39, April 2014.
2. Kimi Kaneko, Daisuke Bekki. “Toward a Discourse Theory for Annotating Causal Relations in Japanese,” In *the 28th Pacific Asia Conference on Language, Information and Computation (PACLIC28)*, pp.460–469, December 2014.
3. Kimi Kaneko, Saku Sugawara, Koji Miheshima, Daisuke Bekki. “Annotation and Analysis of Discourse Relations, Temporal Relations and Multi-Layered Situations Relations in Japanese Texts,” In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pp.10–19, December 2016.

国内会議・研究会

1. 金子 貴美, 戸次 大介. SDRT に基づく日本語談話アノテーションの試み. 情報処理学会 第 214 回 自然言語処理研究会, 2013-NL-214(11), 1-4, November 2013.
2. 金子 貴美, 戸次 大介. SDRT に基づく因果関係認識日本語評価データ構築手法の提案. 言語処理学会 第 20 回年次大会, pp.1079–1082, March 2014.
3. 宇津木 舞香, 稲田 和明, 金子 貴美, 戸次 大介, 乾 健太郎. 形式意味論に基づく出来事間関係認識に向けてリソース構築の展望とテンス「タ」のアノテーション. 言語処理学会 第 21 回年次大会, pp.1036–1039, March 2015.
4. 金子 貴美, 戸次 大介. 因果関係認識のための日本語談話アノテーションとその分析. 第 29 回人工知能学会全国大会, May 2015.