

# 日本語文書内で表現される事象間の 時間的な順序関係の推定

稲田 和明<sup>1,a)</sup> 松林 優一郎<sup>1</sup> 乾 健太郎<sup>1</sup>

受付日 2015年4月30日, 採録日 2015年6月5日

**概要:** 本稿では、文書中の表現間の時間的な順序関係の推定（時間関係認識）について述べる。英語では時間関係認識に関する研究が多数存在するが、日本語では多くなく、特に一般的なドメインの文書を対象とした研究が乏しい。そこで本研究では、日本語における頑健な時間関係認識の確立を目指し、ドメインに依存しない解析手法を分析する。英語では機械学習手法によって成果をあげているが、その知見が日本語でも効果的に働くかは自明ではないため、英語で採用される素性が日本語でも同様に有効かを検証する。また、時間関係認識はまだまだ発展途上であることから、その解析精度の向上を目指し、時間表現に対する依存構造と大規模データから獲得した頻度情報を利用した素性を提案する。さらに、各素性の貢献度、学習曲線などの分析や、実際の解析によって生じた誤り事例の調査を行う。これらの結果、我々が提案した素性によって解析精度の向上が見られたが、大規模データの頻度情報を利用した素性は頻度獲得に改良が必要と分かった。また、英語での手法は日本語でもおおむね有効であったが、言語間の文法的な表現方法の差異によって、日本語では品詞の素性が悪影響を与えていた。学習曲線などの分析から現状では学習データ量が不十分であり、さらなるデータ量増加に価値があること、誤り分析から事象の持つテンス・アスペクトといった機能的意味の解析などが、今後の時間関係認識の発展に必要と分かった。

キーワード：時間情報処理, 自然言語処理, 意味解析

## Temporal Relation Classification between Events Represented in Japanese Texts

KAZUAKI INADA<sup>1,a)</sup> YUICHIRO MATSUBAYASHI<sup>1</sup> KENTARO INUI<sup>1</sup>

Received: April 30, 2015, Accepted: June 5, 2015

**Abstract:** Temporal relation classification (TRC) is the task of classifying temporal relations between two expressions in a given text. While TRC has been widely explored in English, it has rarely been explored in Japanese aside from domain-specific models. In this work, we aim to create a TRC system for Japanese texts that works with general domains. To accomplish this, we first developed a system based on previous machine-learning models in English, and we verified that the same features work for Japanese. We then evaluated novel features that exploit dependency contexts and a co-occurrence frequency obtained by a large-scale corpus. In detail, we evaluated the contribution of each feature, learning curve, and accuracy of each temporal relation type, and we also performed an error analysis. The results show that our proposed features effectively worked, but it is necessary to improve the co-occurrence-based feature. The features we investigated for English are also effective in Japanese; however, part of speech did not work well due to their grammatical differences. The learning curve indicated that the number of training data was insufficient. Finally, in our error analysis, we found that detecting tense, aspect, and modality are necessary for further improvement of our system.

**Keywords:** temporal information processing, natural language processing, semantic analysis

## 1. はじめに

時間的な順序関係の推定（以降、**時間関係認識**と呼ぶ）とは、自然言語文書中の「3/31」「4月中」といった具体的な時を表す時間表現や、「分離した」「合併し」などの事象表現に対して、それらの間の時間的順序を推定することである。図 1 に示す例からは、「3/31」中に「分離」し、「4月中」に「合併」した後「設立」といった時間的な順序関係が読み取れる。このような時間的な順序関係の推定は、事象間の因果関係や文書間の談話的なつながりなどの、自然言語処理の応用的な意味・談話解析において有用と考えられているため、時間情報処理の中でも特に重要視されている問題である。実際に、英語では Chambers らによって時間関係認識を利用した事象表現間の知識ベース [1] が作成され、その知識ベースが共参照解析などの自然言語処理における応用的な解析で効果的に働く [2] など、時間関係認識の有用性が示されている。

現在までに、時間関係認識に関する研究は英語やスペイン語、中国語などの言語では多数報告されている。特に英語では、2 表現間の時間関係をはじめとする時間情報が付与された学習・評価用コーパス TimeBank [3] が早期に開発されたこともあり、評価型ワークショップ TempEval が 3 回も開催される [4], [5], [6] など、さかんに研究されている。このように多数の研究結果が報告されているが、解析精度は F1 値で約 6 割と未解決の問題が残存している。一方日本語では、英語における TimeBank のような時間情報が付与された学習・評価用コーパスが長らく存在しなかったため、時間関係認識に関する研究が乏しかった。料理レシピや大規模な事件・事故・災害などの一部の限定的なドメインを対象とした研究はいくつか報告されている [7], [8] が、それらのドメイン固有の事象間知識や文書構造を解析に利用しており、広いドメインに対する一般性を持った解析手法ではなかった。このようななか、近年 Asahara らによって日本語の時間情報タグ付きコーパス BCCWJ-TimeBank [9] が構築され、日本語でも一般的なドメインの文書において機械学習を用いた時間関係認識とその定量的な評価が可能となり、時間関係認識に関する

研究が報告され始めた [10], [11]。しかし、日本語では時間関係認識に関する研究報告が少ないこともあり、様々な観点からの評価や解析結果に対する分析などの日本語における時間関係認識の地盤となる十分な知見は存在しない。

多数の知見が存在する英語での研究成果は日本語でも有用であると期待できるが、英語と日本語は文法的規則などに多数の差異が存在する距離が離れた言語であるため、それらの知見が日本語でも同様に有効であるかは必ずしも自明ではない。そこで、我々は頑健な時間関係認識の解析手法の確立を目指し、ドメインに依存しない解析手法を対象に詳細な分析を行った。具体的に本研究では、BCCWJ-TimeBank を用いて以下の 4 種類の検証を行った。(1) 英語での基本的な解析手法である機械学習によるアプローチにおいて、各素性の言語依存性および独立性を確かめるために、英語で使用されていた既存の素性を日本語用に整備し、各素性の貢献度の調査や日本語と英語間の比較を行った。(2) 英語でもいまだ発展途上である時間関係認識の解析精度の向上を目指して、我々が新規に提案する「時間表現の係り先の事象表現の周辺情報を利用した素性」と「大規模データから獲得した事象表現対の頻度情報を使用した素性」の有効性を検証した。(3) 日本語の時間関係認識に対する様々な側面からの定量的な評価として、日本語の先行研究との比較や学習曲線の分析などを実施した。(4) 今後の時間関係認識の性能改善のための指針を得るために、著者らが人手で誤り事例をカテゴリに分けて整理し、実際の誤り事例を参照しつつ、それらの解決方法を考察した。

これらの検証の結果、(1) に関しては、英語で使用されている素性の大半は日本語でも同様に有効であることが分かった。ただし、英語における動詞の品詞解析では、過去 (-ed) や進行 (-ing) などのテンス・アスペクトに関わる事象の機能的意味を品詞ラベル上で区別して扱っているが、日本語の品詞体系は一般にこれらの機能的意味を区別しないため、事象が持つ機能的意味の曖昧性をとらえる解析が必要であった。(2) に関しては、時間表現の係り先の事象表現の周辺情報を利用した素性は効果的に働いたが、大規模データから獲得した事象表現対の頻度情報の素性は、統計学的検定で有意性が認められなかった。(3) に関しては、我々のモデルが現在の日本語での最高精度を達成した。その一方で、学習曲線の分析などから、現状では学習データ量が不足していることや、解析対象の表現対の種別によってはモデルの改良が必要であることが分かった。(4) に関しては、事象表現に影響するテンス・アスペクトや否定・仮定・モダリティなどの機能的な表現を持つ意味解析、事象のスクリプト的な知識や上位・下位概念などの事象間知識の拡充とそれらの使用方法の吟味、特定の時間表現が被覆する事象表現の範囲を同定するための文構造解析などが、今後の時間関係認識の発展のために必要であることが

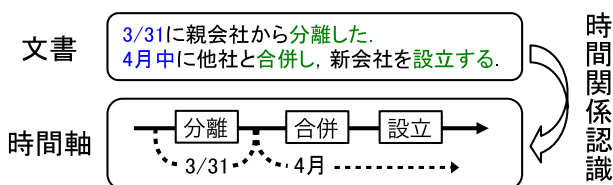


図 1 時間関係認識とは

Fig. 1 Temporal relation classification.

<sup>1</sup> 東北大学大学院情報科学研究科  
Graduate School of Information Sciences, Tohoku University,  
Sendai, Miyagi 980-8579, Japan

a) kazuaki.inada@ecei.tohoku.ac.jp

分かった。

本稿の構成は以下のとおりである。2章では時間関係認識に用いられる時間情報コーパスと時間関係認識の先行研究に関して、英語と日本語の両面から述べる。3章では本研究における時間関係認識の厳密なタスク設定と、評価実験で使用するデータへの前処理に関して説明する。4章では使用した分類手法と採用した素性について説明する。5章では評価実験に関して述べ、実験結果および分析に対する考察を行う。6章は本稿のまとめである。

## 2. 関連研究

本研究で用いる日本語での時間関係認識の学習・評価用コーパスである BCCWJ-TimeBank [9] は、英語の学習・評価用コーパスである TimeBank [3] を参考に作成されており、解析対象の同定や時間関係の表現方法などコーパスの設計に関して様々な類似性が存在する。そのため、本研究では英語における時間関係認識の取り組みを参考に、日本語における時間関係認識のタスクとモデルの設計を行った。具体的には、BCCWJ-TimeBank が作成されて間もないことから、TimeBank が公開された当初の解析器や初期の TempEval を参考に日本語での時間関係ラベルのタスクとモデル構築した。以降の 2.1 節では、TimeBank と BCCWJ-TimeBank の仕様に関して述べ、2.2 節で TimeBank が作成された当初の解析器をはじめ、英語における時間関係認識の取り組みを時系列順に紹介した後、日本語での既存研究について述べる。

### 2.1 TimeBank と BCCWJ-TimeBank

英語では、時間関係認識のための学習・評価用コーパスとして、TimeBank [3] が頻繁に利用されている。TimeBank は、時間表現の抽出に用いられていた既存のアノテーション基準から、時間的な順序関係の付与へ拡張したアノテーション基準 TimeML [12] に従って、時間情報が付与されたコーパスである。TimeBank 以前の研究では、文書中の時間表現の位置と時間表現を正規化した値が付与されたコー

パスを用いて、時間表現の抽出と正規化を行うのみであった [13], [14]。一方 TimeBank では、文書上での時間的な順序関係を扱うために、図 2 のように時間表現と時間的な順序関係を持つと考えられる事象表現の位置を同定したうえで、文書中の 2 つの事象表現間、時間表現と事象表現間、文書作成日時と事象表現間の 3 種類の表現間の時間関係を、Allen の代数論理 [15] に基づく 13 種類のラベルを用いて表現している。

現在 TimeBank は、183 のニュース記事、61,000 語の規模のコーパス TimeBank 1.2.1 が公開されており、評価型ワークショップ TempEval などの時間情報処理に関わる研究で幅広く利用されている。さらに TimeBank を筆頭に、時間情報がより正確に付与されたコーパス、時間関係認識によって自動推定された時間関係を利用したコーパス、英語以外の中国語、スペイン語、イタリア語、フランス語のコーパスなど、様々な用途に対応した時間情報コーパスが作成されたが、長らく日本語では時間関係認識のためのコーパスは開発されなかった。

このようななか、近年日本語においても、TimeML を日本語向けのアノテーション基準に改変し、現代日本語書き言葉均衡コーパス (BCCWJ) [16] へ時間表現や事象表現の位置と、それらの間の時間関係などの時間情報を付与した BCCWJ-TimeBank コーパスが作成された [9]。図 3 に BCCWJ-TimeBank に付与された時間関係の例を示す。BCCWJ-TimeBank では、文書上の時間表現や事象表現の位置を同定したうえで、「E2E: 隣接事象表現間」

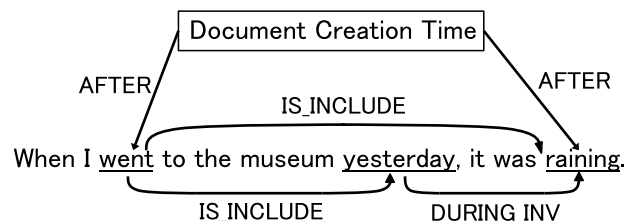


図 2 TimeBank 上の時間関係

Fig. 2 Temporal relations on TimeBank.

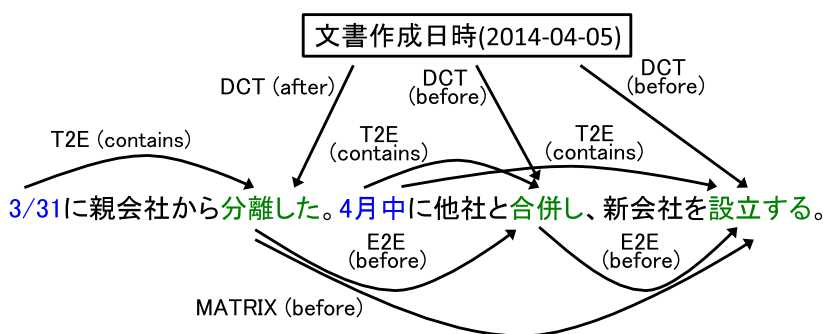


図 3 BCCWJ-TimeBank の概観

Fig. 3 Overview of BCCWJ-TimeBank.

**時間関係ラベルの意味**  
 after(X→Y): XがYよりも後に起こる  
 before(X→Y): XがYよりも先に起こる  
 XとYに時間的な重なりがあり、  
 contains(X→Y): Xの開始点はYの開始点より前  
 Xの終了点はYの終了点より後  
 (\*X:矢印の元、Y:矢印の先)

**時間関係ラベルの付与対象**  
 E2E: 隣接する事象表現間  
 MATRIX: 隣接文の主要事象表現間  
 T2E: 同一文内の時間表現と事象表現間  
 DCT: 文書作成日時と事象表現間



「MATRIX (以降, MAT と略記する): 隣接文の末尾の事象表現間」, 「T2E: 同一文内の時間表現と事象表現間」, 「DCT: 文書作成日時と事象表現間」の4種類の表現間に対して時間関係を示すラベル (時間関係ラベル) が付与されている. なお時間関係ラベルの付与は, 文書作成日時の記載の関係から, BCCWJの全ドメインではなく新聞ドメインのみに限定されている. またE2Eでは, 文書中の隣接事象表現間すべてを付与対象としているため, 同一文内のみではなく文間の事象表現間も付与対象に含まれる点に注意されたい.

表現間の時間関係ラベルとしては, Allenの範囲代数 [15]に基づく13種類のラベルに加え, 部分事象の関係かまったく同一の事象の関係である場合の3種類のラベル (identity, includes, is\_included) と, 時間関係が付与できないことを示す (vague) を加えた計17種類が採用されている. 17種類の時間関係ラベルの概観を図4に示す. たとえば, 時間関係ラベル after(X,Y) は, Yが起きた後にXが起きることを意味し, また met-by(X,Y) は, Yが起きた後その直後に連続してXが起きることを意味している. なお, 時間関係ラベルは作業者3人によって付与されているため, 1つの付与対象に対して3つの時間関係ラベルが記載されている.

BCCWJ-TimeBankには時間関係ラベルのほかに, 時間表現や事象表現へそれらの時間情報を表す属性が人手で付与されている. その中で, 時間関係ラベルの判断に関わる情報を持った属性, value, type, classの3つを紹介する. 時間表現に付与される属性 value は, その時間表現が実際に指し示す値を曖昧性が生じない形に正規化したものである. たとえば「3/31」という時間表現に対しては, 文書作成日時や他の時間表現を手がかりとして「2014-03-31」という値が付与される. type は, 時間表現に付与さ

れる属性であり, 時間表現が示す期間によって区別され, 「DATE, TIME, DURATION, SET」のいずれかが付与されている. また class は事象表現に付与される属性であり, 「OCCURRENCE, REPORTING, PERCEPTION, ASPECTUAL, LACTION, LSTATE, STATE」の7種類が付与されている. これらは, 事象のテンス・アスペクトの体系に関する言語学の文献 [17] の分類をもとに作成されており, 動詞の語彙の意味を考慮した時間的性質に関わる分類である. 表1に type および class に与えられる値とその意味を示す. なお, class などの属性をはじめとする BCCWJ-TimeBank の詳細な情報に関しては, 参考文献 [9], [18], [19] を参照されたい.

2.2 時間関係認識

英語圏では, TimeBank が作成された初期のころ, 入力として与えられた時間表現や事象表現に対して, それらに間に付与された時間関係ラベルを, 最大エントロピー法やサポートベクタマシン (SVM) によって分類する研究が行われていた [20], [21]. 当時は機械学習で使用する素性も, TimeBank に付与された時間情報や入力表現対の表層・品詞といった形態素情報などの単純なものであった.

その後, 初期の研究結果をもとにコーパスの整備やタスクの設計が行われると, 時間表現や事象表現間の時間的順序関係の推定を目的とした評価型ワークショップ TempEval\*1が開催された [4], [5], [6].

初期の TempEval-1 では, 「(1) 同一文内の時間表現と事象表現間, (2) 文書作成日時と事象表現間, (3) 隣接文の主要事象表現間」の3種類の表現間に対して, それらに間に付与された時間関係ラベルを推定するタスクが設けられた. 問題の簡単化のため, 時間関係ラベルを13種類から主要な6種類にグループ化することで, 解析難易度を下げるといった調整も行われた. SVMなどの機械学習による分類

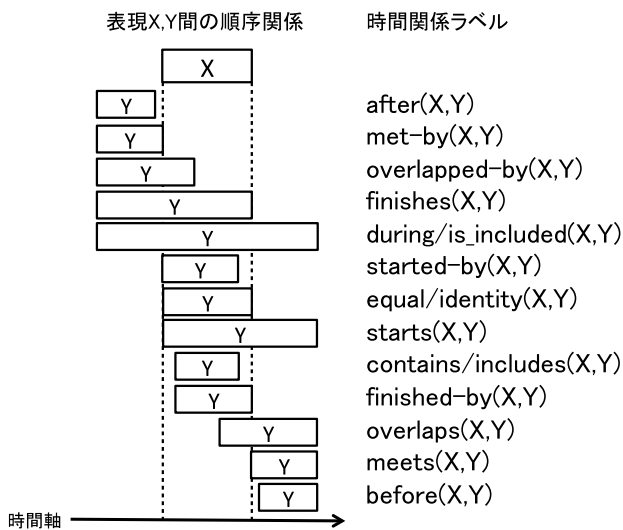


図4 時間関係ラベルが意味する表現 X と表現 Y 間の時間順序  
Fig. 4 Temporal orders between X and Y represented by temporal relation labels in BCCWJ-TimeBank.

表1 BCCWJ-TimeBankにおける付加属性  
Table 1 Attributes annotated on BCCWJ-TimeBank.

属性	属性の値	意味
type	DATE	日付を示す時間表現
	TIME	時刻を示す時間表現
	DURATION	期間を示す時間表現
	SET	頻度や集合を示す時間表現
class	OCCURRENCE	事象表現一般
	REPORTING	表現活動動詞
	PERCEPTION	認知・知覚動詞
	ASPECTUAL	アスペクト動詞
	LACTION	内包的な動作
	LSTATE	内包的な静態動詞
	STATE	静態動詞, 形容詞, 形状詞

\*1 本稿では, 文献 [4] を TempEval-1, 文献 [5] を TempEval-2, 文献 [6] を TempEval-3 と呼称する.

モデル以外にも、時間関係認識を系列ラベリング問題としてモデル化したシステムや [22]、形態素情報や TimeBank に付与された時間情報の組み合わせ規則で時間関係ラベルを推定するルールベースのシステム [23] など、様々な手法が提案された。

次の TempEval-2 では、TempEval-1 でのタスクに加え、同一文内の事象表現間を対象とした時間関係認識のタスクや、時間表現や事象表現の抽出、時間表現の正規化などの時間情報処理が行われた。英語のみではなく、新たに作成されたスペイン語のコーパスを用いて、スペイン語での時間関係認識も実施された [24]。このころから、文全体での時間関係の最適化を行うために Markov Logic や条件付き確率場などの複雑な機械学習手法が利用され [25], [26]、また文書中から取得できない事象表現に関する語彙的な知識を、WordNet<sup>\*2</sup>などの外部リソースからの取得を試みるシステムが提案され始めた [27]。

最新の TempEval-3 では、6 種類に制限していた時間関係ラベルを 13 種類に戻したうえで時間関係認識が行われた。時間関係ラベルの種類数が多く、その分布に大きな偏りが存在する困難なラベル分類問題であったが、与えられた表現間の時間関係ラベルを推定するタスクにおける最高精度のシステムは約 56% の F1 値を達成した。さらに、与えられた表現間に対する時間関係認識だけでなく、平文からその文書内の時間的な順序関係を推定することを目的に、時間表現や事象表現の抽出の後、それらの表現間の時間関係を推定するという時間情報処理を総合的に行う挑戦的なタスクも設けられた。

また、英語では TempEval 以外にもさかんに時間関係認識の研究が報告されており、その中でも本研究の取り組みと類似した研究をあげる。D'Souza ら [28] は、類義語や対義語、上位・下位語などの事象知識、主語や目的語などの意味役割や文間の談話関係などの解析器の出力を素性として利用したシステムを提案した。これらの素性を、時間関係認識で一般的に使用されている素性と多数の規則によって組み合わせることで、高い効果を引き出すことに成功している。また、Mirza ら [29] は TempEval で採用されている基本的な素性に関して、これらの素性の貢献度を詳細に解析し、その調査結果から有用な素性だけに絞ったシステムを構築することで、TempEval-3 に投稿された最高精度のシステムと同等の精度を達成した。これらの研究では、評価実験や分析の乏しい TempEval とは異なり、時間関係ラベルごとの推定精度や使用素性の貢献度に対する検証などの細かな定量的評価を実施している。我々もこれらの研究にならい、システム全体の精度のみではなく、様々な観点から定量的な評価を行った。

このように、英語では時間関係認識に関する研究が多数

実施されている。しかしその一方で、いまだ解決できていない問題が存在するにもかかわらず、人手で実際の事例を見るなどして、解析の際に生じた現象や解析に必要な手がかりをとらえるなど、時間関係認識の解析結果に対する調査・分析は十分に報告されていない。

日本語でも、拡張固有表現抽出における時間表現の抽出 [16]、述語項構造解析における事象表現の抽出 [30]、拡張モダリティ解析における事象表現の時制の解析 [31] など、自然言語処理における情報抽出や意味解析の研究成果の一部として時間情報処理に該当する研究は存在するが、文書中の表現間を対象とした時間関係認識の研究は多くはなかった。このようななか、少量ではあるが日本語で報告されている時間関係認識の研究において、本研究のタスク設定に近い研究を紹介する。

まず島崎らは、レシピテキスト上で事象表現間の時間関係認識を実施している [7]。彼らは、時間関係ラベルの推定にレシピテキストのドメインに特化した事象間知識を与えることで、どの程度まで推定精度を高められるかを検証した。しかし、彼らの手法をレシピテキスト以外のドメインへ適用するためには、適用先のドメインに応じた事象間知識が必要となる。さらに、あらゆるドメインで時間関係認識に効果的な事象間知識を作成できる保証はないため、様々なドメインの文書に対する頑健性は乏しい。また吉川らは、我々と同様に BCCWJ-TimeBank 上で時間関係認識を実施している [10]。彼らは定量的な評価として、TempEval に投稿されたシステムとの比較を行っているが、システム全体での検証のみにとどまっており、有効な素性の吟味や学習曲線の分析といった、今後の日本語における時間関係認識の発展に有用となる調査が十分に行われているとはいえない。また英語と同様に、実際の解析結果に対する分析も報告されていない。

### 3. 時間関係ラベルの付与

#### 3.1 時間関係認識タスク

本研究では、英語での先行研究の 1 つである TempEval-2 [5] のタスク設定を参考に、日本語での時間関係認識のタスクを構築した。

TempEval-2 の時間関係認識タスクでは、入力として「(1) 解析対象の表現対、(2) それらが含まれる文、(3) TimeBank に付与された value をはじめとするコーパスの付加属性」の 3 つが与えられ、(1) の表現対に付与された時間関係ラベルを出力するタスクであったが、解析対象となる入力表現対の種類によって 4 つのタスクに分けられていた。TempEval-2 における入力表現対の種類とタスク分類を表 2 に示す。なお、BCCWJ-TimeBank における時間関係ラベルの付与対象 (E2E, MAT, T2E, DCT) は、TempEval-2 のタスクと対応関係が存在するため、表 2 にあわせて記載した。ただし、TempEval-2 の Task-C は直接係

\*2 <http://wordnet.princeton.edu/>

表 2 TempEval-2 のタスク設定と BCCWJ-TimeBank との対応  
**Table 2** Tasks in TempEval-2 and correspondence to BCCWJ-TimeBank.

Task	解析対象の表現対	日本語との対応
C	同一文内の時間表現と事象表現間	T2E
D	文書作成日時と事象表現間	DCT
E	隣接文間の主節の事象表現間	MAT
F	構文的支配関係を持つ事象表現間	E2E

り受け関係にある時間表現と事象表現間を解析対象としているが、BCCWJ-TimeBank の T2E では同一文内すべての時間表現と事象表現間を時間関係ラベルの付与対象としているため、TempEval-2 の Task-C と BCCWJ-TimeBank の T2E 間には多少の差異が存在する。また、TempEval-2 の Task-F では文内の事象表現間を解析対象としているが、BCCWJ-TimeBank の E2E では文間の事象表現間も時間関係ラベルの付与対象に含まれるため、TempEval-2 の Task-F と BCCWJ-TimeBank の E2E 間には大きな差異が存在することに注意されたい。

本研究では、BCCWJ-TimeBank 上で時間関係ラベルが付与された表現対の種別をもとに、上述の TempEval-2 のタスク設定と類似させた 4 つのタスク (E2E, MAT, T2E, DCT) を設ける。以降では、各タスクにおける厳密な定義を述べる。なお、時間関係ラベルや使用するデータに対して制限を加えるが、各タスクの定義とは独立しているため 3.2 節で述べる。

### 3.1.1 Task-E2E

BCCWJ-TimeBank における E2E の付与対象は、前述のとおり文内と文間の事象表現間の両方が含まれるが、本研究では E2E の解析対象を文内の問題のみに限定する。自然言語処理分野の研究の 1 つである述語項構造解析でも、時間関係認識と同様に文内と文間の解析対象が存在するが、これらは問題の性質が異なるため、問題を分離したうえで取り組む研究が多い。また TempEval-2 では、問題の単純化と問題の性質の差異から、文内と文間の時間関係認識を別々のタスクに分けて扱っていた。しかし BCCWJ-TimeBank の E2E では、文内や文間に関係なく文書中の隣接する事象表現間に対して時間関係ラベルが付与されているため、そのままでは文内と文間の問題が混在してしまう。そこで本研究でもこれらの先行研究にならない、E2E では解析対象を文内に限定することで、問題の切り分けと単純化を図る。

上述の制約のもと、解析対象となる事象表現対において、文書上で先に出現する事象表現を  $w_1$ 、後に出現する事象表現を  $w_2$  とする。なお、 $w_1$  と  $w_2$  は、それぞれの表現の表層に加え、BCCWJ-TimeBank に人手で付与された属性を含むものとする。また、 $w_1$  と  $w_2$  が含まれる文を  $s$ 、 $w_1$  と  $w_2$  間に付与された時間関係ラベルを  $r$  とし、 $w_1$  と  $w_2$  間の時間関係を  $r(w_1, w_2)$  と表す。このとき、E2E の時間

関係認識タスクは、入力 ( $w_1, w_2, s$ ) に対して、 $r(w_1, w_2)$  の  $r$  を出力とする問題として与えられる。

### 3.1.2 Task-MAT

MAT では、解析対象となる事象表現対において、文書上で先に出現する事象表現を  $w_1$ 、後に出現する事象表現を  $w_2$  とする。なお、 $w_1$  と  $w_2$  は、それぞれの表現の表層に加え、BCCWJ-TimeBank に人手で付与された属性を含むものとする。また、 $w_1$  が含まれる文を  $s_1$  とし、 $w_2$  が含まれる文を  $s_2$ 、 $w_1$  と  $w_2$  間に付与された時間関係ラベルを  $r$  とし、 $w_1$  と  $w_2$  間の時間関係を  $r(w_1, w_2)$  と表す。ただし以降では、便宜上  $s_1$  と  $s_2$  を合わせて  $s$  と記述する。このとき、MAT の時間関係認識タスクは、入力 ( $w_1, w_2, s$ ) に対して、 $r(w_1, w_2)$  の  $r$  を出力とする問題として与えられる。

### 3.1.3 Task-T2E

T2E では、解析対象となる時間表現と事象表現間において、時間表現を  $w_1$ 、事象表現を  $w_2$  とする。なお、 $w_1$  と  $w_2$  は、それぞれの表現の表層に加え、BCCWJ-TimeBank に人手で付与された属性を含むものとする。また、 $w_1$  と  $w_2$  が含まれる文を  $s$ 、 $w_1$  と  $w_2$  間に付与された時間関係ラベルを  $r$  とし、 $w_1$  と  $w_2$  間の時間関係を  $r(w_1, w_2)$  と表す。このとき、T2E の時間関係認識タスクは、入力 ( $w_1, w_2, s$ ) に対して、 $r(w_1, w_2)$  の  $r$  を出力とする問題として与えられる。

### 3.1.4 Task-DCT

DCT では、解析対象となる文書作成日時と事象表現間において、文書作成日時を  $w_1$ 、事象表現を  $w_2$  とする。なお、 $w_1$  と  $w_2$  は、それぞれの表現の表層に加え、BCCWJ-TimeBank に人手で付与された属性を含むものとする。また、事象表現  $w_2$  が含まれる文を  $s$ 、 $w_1$  と  $w_2$  間に付与された時間関係ラベルを  $r$  とし、 $w_1$  と  $w_2$  間の時間関係を  $r(w_1, w_2)$  と表す。このとき、DCT の時間関係認識タスクは、入力 ( $w_1, w_2, s$ ) に対して、 $r(w_1, w_2)$  の  $r$  を出力とする問題として与えられる。

## 3.2 データセット

本研究では、3.1 節で定義した 4 つのタスクに対して、表 3 に示す BCCWJ-TimeBank のデータを直接利用のではなく、以下の 2 つの処理を加えた。

### (1) 時間関係ラベルのグループ化

BCCWJ-TimeBank では計 17 種類の時間関係ラベルが採用されているが、表 3 に示すとおり一部の時間関係ラベルはコーパス中にごく少数しか存在しないため、少数派の時間関係ラベルの推定は困難となる。実際に、後述する 5.4 節における学習曲線の分析からも推定の困難さを確認している。そこで、TempEval-1 や TempEval-2 の実験設定にならない [4], [5]、表 4 の変換表を用いて時間関係ラベルを機械的に変換することで、時間関係ラベルの種類数を 6



表 3 BCCWJ-TimeBank の時間関係ラベル量

Table 3 The number of label instances on BCCWJ-TimeBank.

	E2E	MAT	T2E	DCT
after	432	198	315	1961
met-by	2	1	2	0
overlapped-by	0	0	20	2
finished	1	0	0	0
during	67	43	62	217
started-by	2	0	0	0
equal	154	29	19	0
starts	0	0	2	0
contains	117	64	671	63
finished-by	0	0	0	0
overlaps	0	0	32	1
meets	22	0	2	0
before	713	294	288	572
is_included	1	0	0	0
identity	2	2	0	0
includes	1	0	0	0
vague	128	48	100	38
合計	1642	679	1513	2854

表 4 時間関係ラベルのグループ化

Table 4 Grouping of temporal relation labels.

グループ化後	グループ化前
AFTER	after, met-by
BEFORE	before, meets
B-or-O	starts, overlaps
O-or-A	overlapped-by, finishes
OVERLAP	during, started-by, equal, contains, finished-by, is_included, identity, includes
VAGUE	vague

表 5 本研究で使用した時間関係ラベルのデータ量

Table 5 The number of grouped label instances.

	E2E	MAT	T2E	DCT
AFTER	177	201	323	1961
BEFORE	582	305	303	581
B-or-O	0	0	43	1
O-or-A	1	0	22	2
OVERLAP	351	222	814	290
VAGUE	27	48	100	38
合計	1138	776	1605	2873

種類に簡略化して扱う。

(2) 作業員間で付与が一致したデータに限定

2.1 節で述べたとおり、BCCWJ-TimeBank では 3 人の作業員によって時間関係ラベルを付与している。本研究では、作業員 3 人の時間関係ラベルの付与が一致した表現対のみに解析対象を限定した。なお、時間関係ラベルのグループ化後に作業員間の付与が一致しているかを判定した。

上述の (1) と (2) の処理の結果、データ量は表 5 と

なった。本研究ではこの表 5 のデータを用いて評価実験を行った。

4. 分類手法

4.1 統計モデル

本研究では、入力として与えられた表現対  $w1$  と  $w2$  間の時間関係ラベル  $r$  の推定に最大エントロピー法を採用した。最大エントロピー法は多クラス分類問題でよく利用されるモデルであり、分類器に組み込む特性を素性として容易に記述できるため、再現が容易であることが知られている。

多クラス分類問題における最大エントロピー法では、入力  $(w1, w2, s)$  から得られる素性ベクトル  $f$  を用いて、6 種類の時間関係ラベルごとに式 (1) で与えられる条件付き確率を求め、その確率が最大となる時間関係ラベル  $r$  を推定値として出力する。

$$p(r|w1, w2, s) = \frac{1}{Z} \exp(\mathbf{w}_r \cdot \mathbf{f}(w1, w2, s)) \quad (1)$$

ここで  $\mathbf{w}$  は入力の素性ベクトル  $f$  と時間関係ラベル  $r$  に対する重みベクトルであり、学習では教師データ中の尤度を最大にするような  $\mathbf{w}$  を求める。  $Z$  は正規化項であり、以下の式 (2) で表される。

$$Z = \sum_r \exp(\mathbf{w}_r \cdot \mathbf{f}(w1, w2, s)) \quad (2)$$

4.2 使用素性

本研究で採用した素性の一覧を表 6 に示す。英語における代表的な解析器 [5], [6], [28], [29] で採用されている素性のうち、日本語においても効果が期待できそうな素性を日本語用に読み替えたり、変更したりしたものを導入した。ただし、日本語では使用できるデータ量が乏しいため、疎データ問題が発生しにくい基本的な素性のみ限定した。さらに本研究独自の素性として、時間表現の係り先の事象表現の周辺情報をとらえる素性と、大規模データから獲得した事象表現対の頻度情報を利用した素性を提案する。以降、表 6 に示した各素性の詳細に関してカテゴリごとに説明する。なお各素性の作成において、形態素や品詞の解析に IPA 品詞体系の NAIST Japanese Dictionary 0.6.3b<sup>\*3</sup>を辞書として用いた MeCab 0.996<sup>\*4</sup>、文節および統語構造の解析に CaboCha 0.68<sup>\*5</sup>を使用した。

4.2.1 形態素

形態素に関する素性として、IPA 品詞体系における基本形 (Base)、第 1 階層の品詞 (POS)、第 2 階層の品詞 (POS1)、活用形 (ConjForm) を採用した。品詞は、英語で採用されている Penn TreeBank<sup>\*6</sup>の品詞体系と同程度の粒度となる第 2 階層の品詞まで採用した。また活用形は、

\*3 <http://sourceforge.jp/projects/naist-jdic/>

\*4 <http://taku910.github.io/mecab/>

\*5 <http://taku910.github.io/cabocha/>

\*6 <https://www.cis.upenn.edu/~treebank/>

表 6 時間関係ラベルの推定に使用した素性  
Table 6 Features for temporal relation classification.

カテゴリ	素性名	詳細分類	説明	E2E	MAT	T2E	DCT
形態素	Base	event	事象表現自身と前後 2 形態素の基本形	✓	✓	✓	✓
		time	時間表現自身と前後 2 形態素の基本形			✓	
		combi	表現対間の基本形の組み合わせ	✓	✓	✓	
	POS	event	事象表現自身と前後 2 形態素の第 1 階層の品詞	✓	✓	✓	✓
		time	時間表現自身と前後 2 形態素の第 1 階層の品詞			✓	
		combi	表現対間の第 1 階層の品詞の組み合わせ	✓	✓	✓	
	POS1	event	事象表現自身と前後 2 形態素の第 2 階層の品詞	✓	✓	✓	✓
		time	時間表現自身と前後 2 形態素の第 2 階層の品詞			✓	
		combi	表現対間の第 2 階層の品詞の組み合わせ	✓	✓	✓	
	ConjForm	event	事象表現自身と前後 2 形態素の活用形	✓	✓	✓	✓
		time	時間表現自身と前後 2 形態素の活用形			✓	
		combi	表現対間の活用形の組み合わせ	✓	✓	✓	
BCCWJ-TimeBank の付加属性	Value		BCCWJ-TimeBank の付加属性 value			✓	✓
	TypeClass	type	BCCWJ-TimeBank の付加属性 type			✓	✓
		class	BCCWJ-TimeBank の付加属性 class	✓	✓	✓	✓
		combi	表現対間の type や class の組み合わせ	✓	✓	✓	✓
文構造	DepDistance		対象表現対間の係り受け距離	✓		✓	
	Order		対象表現のどちらが先に出現するか?			✓	
	ExistTime		対象表現対間に時間表現が存在するか?	✓	✓	✓	
	HeadTail	event	事象表現が文頭もしくは文末の語か?	✓	✓	✓	✓
time		時間表現が文頭もしくは文末の語か?			✓		
combi		表現対間の文頭・文末の組み合わせ	✓	✓	✓		
機能表現	FuncExpr	event	事象表現に付随する助詞や助動詞, 非自立語	✓	✓	✓	✓
		time	時間表現に付随する助詞や助動詞, 非自立語			✓	
		combi	表現対間の機能表現の組み合わせ	✓	✓	✓	
		pattern	事象表現に付随する特定の助詞や助動詞, 非自立語	✓	✓	✓	✓
拡張モダリティ	Modality	event	事象表現の拡張モダリティ	✓	✓	✓	✓
		combi	表現対間の拡張モダリティの組み合わせ	✓	✓		
係り先の事象表現	DepEvent		時間表現の係り先の事象表現の文脈情報			✓	
大規模データの頻度情報	EventsCount		大規模データ上の事象表現対の頻度情報	✓	✓		

次の形態素の表現によってその値が変化するため、助動詞「た」などの時間関係に大きく影響を与える表現を間接的にとらえることができると考えて導入した。基本形, 第 1 階層の品詞, 第 2 階層の品詞, 活用形のそれぞれで, 解析対象の表現対の形態素自身とそれらの周辺の前後 2 形態素の値を使用したほか, 解析対象の表現対の形態素情報を組み合わせた値を素性として用いた。

#### 4.2.2 BCCWJ-TimeBank の付加属性

BCCWJ-TimeBank で時間表現や事象表現に人手で付与されている属性のうち, 時間表現を正規化した値である value (Value) と, 時間表現および事象表現を分類した値である type と class (TypeClass) を素性として用いた。また組合せ素性として, E2E と MAT では解析対象の事象表現対の class を組み合わせた値, T2E と DCT では解析対象の時間表現の type と事象表現の class を組み合わせた値を採用した。

#### 4.2.3 文構造

英語では, 深い構文構造をとらえる統語情報解析器 Enju<sup>\*7</sup> の出力を素性として取り入れたシステムが高い精度を達成している [32]。日本語においても, 解析対象の表現対の文構造が時間関係認識に有効かを確かめるため, 以下の 4 つの素性を採用した。

- (1) 解析対象の表現間の係り受け距離 (DepDistance)
- (2) 時間表現と事象表現のどちらが文書上で先に出現するか (Order)
- (3) 表現対間に時間表現が存在するか (ExistTime)
- (4) 解析対象の表現が文頭の文節に含まれるか, もしくは文末の文節に含まれるか (HeadTail)

Order は, コーパスの制約上, 解析対象の表現対のうち, どちらが文書上で先に出現するかが決まっていない T2E を対象とした素性であり, その時間表現と事象表現の順序を判定するために導入した。また, ExistTime における時

\*7 <http://www.nactem.ac.uk/enju/>



間表現の判定には、CaboChaによる固有表現抽出の結果を使用した。HeadTailでは、解析対象の表現それぞれの文頭・文末判定のみではなく、解析対象の表現間の文頭・文末判定を組み合わせた素性も採用した。具体的には、解析対象の各表現を部分集合、各文頭と文末判定の素性をノードと見なしたとき、 $K_{2,2}$ の完全2部グラフとなるノード間を組み合わせた値を素性として採用した。

#### 4.2.4 機能表現

英語では事象表現に関する機能的意味として、TimeBankに付与されているテンスやアスペクトなどの付加属性を素性として利用しているが、BCCWJ-TimeBankにはテンス・アスペクトに相当する時間情報は付与されていない。そこで、これらの素性の代わりとして、時間表現や事象表現に付随する機能的な表現（以降、機能表現と呼ぶ）が日本語におけるテンス・アスペクトの役割を担うと考え、日本語独自の新規素性として採用した（FuncExpr）。本研究では、時間表現や事象表現と同一の文節内の助詞や助動詞、非自立的な動詞を機能表現と見なし、その表層を素性として用いた。さらに、解析対象の表現対がどちらも事象表現であるE2EとMATでは、解析対象の表現対それぞれの機能表現の表層を組み合わせた値も素性として導入した。

また、日本語における時間関係認識の先行研究の1つである吉川らの研究では、解析対象となる事象表現が「ている」などの特徴的な語尾を持つかを判断する素性を採用している。そこで我々も、時間関係認識に深く関わると考えられる機能的な表現を判断するために、特定の表現のパターン一致による真偽値の素性を採用した。具体的には、「た、が」がそれぞれ事象表現が含まれる文節の末尾に存在するか、「ている、ていた、ており、ながら」がそれぞれ事象表現と同じ文節に存在するかを判断する素性を導入した。

#### 4.2.5 拡張モダリティ

拡張モダリティとは、書き手が表明している真偽判断や価値判断といったモダリティ情報に、時制や肯否などの事象情報を統合した意味情報体系である[33]。この拡張モダリティの中で、時間関係と関連を持つと想定される「時制、真偽判断、仮想」の3つのタグを素性として用いた（Modality）。時制タグは、英語圏で採用されている素性「TimeBankでの付加属性 tense」、真偽判断タグは「TimeBankでの付加属性 polarity」の代用品として採用した。また仮想タグは、事象表現が仮想世界での出来事であるかどうかによって、解析対象の表現間に時間的な順序関係を考慮できるかを判断するために採用した。

使用した拡張モダリティタグの体系は以下のとおりである。時制タグは事象表現と発話時との相対時を示す「未来、非未来」の2種類、真偽判断タグは事象表現の成立・不成立を「成立、不成立、不成立から成立、成立から不成立、高確率、低確率、低確率から高確率、高確率から低確率、0」の9種類、仮想タグは仮定された条件の有無を示

す「条件、帰結、0」の3種類で表現している。拡張モダリティタグとしてBCCWJに人手で付与されている値を利用することも考えたが、BCCWJ-TimeBankの時間関係ラベルが付与された事象表現と一致する部分が少数であったため、BCCWJに付与された拡張モダリティタグの値は使用せず、水野らの自動解析器[34]の出力を拡張モダリティの素性に用いた。なお、解析対象の表現がどちらも事象表現であるE2EとMATでは、同種の拡張モダリティタグの値を組み合わせた素性も導入した。

#### 4.2.6 係り受け先の事象表現

時間表現には「日、時、分」などの特定の表現と数値を組み合わせた定型的な表現が多数存在するため、時間表現やその周辺の語から時間関係ラベルを判断するための材料を取得しにくいと考えられる。そこで我々は、不足するであろう判断材料の補填のため、時間表現の係り先の事象表現の周辺文脈を素性として導入すること新たに提案する。たとえば、「12日に大会の日程を発表したが、取り消すことになった」という文において、「12日」と「取り消す」の間の時間関係を考える際、「12日」の係り先の事象表現にあたる「発表した」の周辺文脈を素性として取り入れることで、時間表現に不足しがちな判断材料を補う。我々は、時間表現の係り先の事象表現の文脈情報として、係り先の事象表現とその周辺の2形態素の基本形、係り先の事象表現との依存構造木上の距離、係り先の事象表現に付随する機能的な表現を素性として採用した（DepEvent）。

#### 4.2.7 外部知識・大規模データの頻度情報

我々は、一部の事象間には因果関係や手続き的な動作順序などのように、各事象が成立する時間的な順序関係に偏りが成立していると仮定し、これを素性として利用することを考えた。具体的には、「寝る」と「消す」の間には、「(電気などを消した後に寝る)のように「消す」から「寝る」方向に時間的な順序関係の偏りが存在している」といった考えである。一般に、事象間知識を素性として導入する際には、既存の知識ベースがよく利用される。時間関係認識のタスクでも、ChengらがVerbOcean<sup>\*8</sup>の事象間知識の導入を試みたが、精度向上が見られなかったと報告している[22]。これは、既存の知識ベースでは実世界で記述される事象表現対に対する被覆率が低く、解析対象に適用した際、十分にヒットしないことが原因として考えられる。実際に、我々が日本語WordNet<sup>\*9</sup>やALAGIN動詞含意関係データベース<sup>\*10</sup>などの事象間知識に対して、BCCWJ-TimeBankの時間関係が付与された事象表現対に適用したところ、その大半がヒットしなかった。そこで、本研究では事象間知識の被覆率を高めるために、手がかり

\*8 <http://demo.patrickpantel.com/demos/verboccean/>

\*9 <http://compling.hss.ntu.edu.sg/wnja/>

\*10 <https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html>

表 7 大規模データからの頻度抽出に用いた手がかり表現

Table 7 Clue expressions for a temporal order of two event expressions.

BEFORE	AFTER
後, 結果, 後々, 今後, その後, 直後, 以降, すぐ, 将来, 未来	前, 前もって, 直前, 以前, この前, 先, 先立つ, あらかじめ

表現を用いた因果関係抽出の手法 [35] を参考として, Web 上の大規模データから事象間知識の獲得を試みた.

我々は, 時間関係ラベル BEFORE と AFTER に対して, それぞれ時間的な順序関係を判断する手がかりとなるフレーズを表 7 のように定めた. その後, 統語係り受け上で, 事象表現 X と Y が表 7 の手がかり表現の親と子にあたるパターン (X → 手がかり表現 → Y) の頻度を, Web 文書約 10 億文から獲得した. そして, 獲得した頻度をもとに式 (3), (4), (5) を用いて事象表現間の時間関係の偏り Temporal Bias (以降, TB と略す) を求めた.

$$B\_Bias(X, Y) = Before(X, Y) + After(Y, X) \quad (3)$$

$$A\_Bias(X, Y) = Before(Y, X) + After(X, Y) \quad (4)$$

$$TB(X, Y) = B\_Bias(X, Y) - A\_Bias(X, Y) \quad (5)$$

頻度カウントの際は, 動詞もしくは形容詞を含む文節の内容語を事象表現として扱った. また, 式 (3), (4) の Before(X, Y) および After(X, Y) は, それぞれ表 7 の BEFORE および AFTER の手がかり表現で獲得した事象表現対 (X, Y) の頻度であり, Before(Y, X) および After(Y, X) は同一の事象表現対 X と Y において, 統語係り受けパターンの親と子を入れ替えた際の頻度である. さらに, B.Bias および A.Bias は, X と Y 間の BEFORE および AFTER の偏りを示す値であり, 数値が高いほどその事象間に BEFORE もしくは AFTER の偏りが存在していること意味する. この B.Bias と A.Bias を統合した指標 TB を素性として用いるが, TB はとりうる値の最大値と最小値に大きな開きがあり, 値が疎になることが予想されたため, 以下の式を用いて, TB を対数で抑え,  $-5 \sim 5$  の整数値に制限した値を素性として使用した (EventsCount). 実際に EventsCount の素性を BCCWJ-TimeBank の事象表現対に適用したところ, ヒット件数は全体の約 7 割であった.

$$\begin{cases} 5 & (TB > 0 \wedge \log_5 TB \geq 5) \\ \lfloor \log_5 TB \rfloor & (TB > 0 \wedge 0 < \log_5 TB < 5) \\ -\lfloor \log_5 (-TB) \rfloor & (TB < 0 \wedge 0 < \log_5 (-TB) < 5) \\ -5 & (TB < 0 \wedge \log_5 (-TB) \geq 5) \\ 0 & (\text{else}) \end{cases}$$

## 5. 評価実験

本稿では, 日本語における時間関係認識の評価実験として 6 つの実験を実施した. 5.1 節では, 先行研究を参考に作

成した既存素性と我々が新規に提案した素性が, それぞれ時間関係認識にどの程度貢献するかを調査した. さらに, 日本語と英語の間で有効な素性に差異が存在するかを確認するため, 5.2 節で本研究で使用した既存素性の貢献度と英語での貢献度調査との比較を実施した. また様々な観点からの定量的評価として, 5.3 節では我々のモデルと先行研究間の正答率の比較, 5.4 節では現状のコーパス量が適切であるかを測るための学習曲線の分析, 5.5 節では時間関係ラベルごとの F1 値の調査を実施した. そして, 実際の推定結果への分析として, 人手で実施した誤り事例に対する調査結果を 5.6 節で報告する.

実験では, 表 5 に示すデータを学習・推定に使用し, 5 分割交叉検定による評価を行った. なお, 重みベクトルの学習には L2 正則化による L-BFGS 法を用い, モデルの学習と時間関係ラベルの推定には Classias [36] を利用した.

### 5.1 各素性の貢献度

日本語での時間関係認識において, 英語の先行研究をもとに作成した既存素性と, 我々が新たに提案した素性がどの程度時間関係ラベルの推定に貢献しているかを確かめた.

#### 5.1.1 形態素カテゴリの素性の貢献度

調査対象の各素性に対する検証を実施する前に, 最も基本的な素性である形態素に関する素性の効果を確認するため, 表 6 の形態素に属するカテゴリに限定して素性の貢献度調査を実施した. 表 8 は, すべての素性を採用したモデル (Full) から, 表 6 で形態素のカテゴリに属する素性を素性名単位で 1 つ取り除いたときの正答率である\*11.

表 8 より, 第 1 階層の品詞 (POS) や第 2 階層の品詞 (POS1) はすべてのタスクで正答率の向上に寄与していないことが確認できる. したがって, 日本語では粒度にかかわらず, 品詞情報に関する素性は時間関係の推定に有効でないことが分かった. また活用形 (ConjForm) は, DCT のみで正答率が向上し, さらに統計学的検定でも有意性を示した. しかし, E2E, MAT, T2E では効果が見られなかったことから, 文書内の 2 つの表現間を解析対象とするタスクでは十分に働かないことが分かった.

表 8 の結果を受け, 5.1.2 項の実験では, POS と POS1 は採用せず, また ConjForm は DCT のみで採用を考慮することとする.

#### 5.1.2 提案した各素性の貢献度

表 9 は, 表 8 で悪影響を与えていた素性を除く表 6 のすべての素性を使用したモデル (Full) から, 表 6 に記し

\*11 捕捉実験として, MeCab に JUMAN 品詞体系の辞書 (JUMAN 辞書 7.0) を用いて得た品詞, 品詞細分類, 活用形を, 表 6 の POS, POS1, ConjForm の作成に利用したモデルで検証を実施したところ, 表 8 に示す IPA 品詞体系での統計学的検定と同様に, DCT における「ConjForm」のみで有意性が認められた. よって本稿では, IPA 品詞体系と JUMAN 品詞体系に大きな差異が存在しないと判断し, IPA 品詞体系を採用した検証結果のみを記載する.

表 8 形態素の素性に関する貢献度  
Table 8 Impacts of morphological features.

	E2E		MAT		T2E		DCT	
	正答率	-Δ	正答率	-Δ	正答率	-Δ	正答率	-Δ
全素性使用 (Full)	65.90		48.32		68.72		80.54	
Full-Base	63.53	-2.37*	47.55	-0.77	63.61	-5.11*	79.64	-0.90
Full-POS	66.78	0.88	49.09	0.77	69.22	0.50	80.79	0.25
Full-POS1	66.43	0.53	48.32	0.00	69.72	1.00	80.82	0.28
Full-ConjForm	66.25	0.35	49.10	0.78	68.72	0.00	79.39	-1.15*

\* Full を比較対象とした 5-fold cross-validated paired t test による両側検定で、有意水準 0.05%以上の有意性を示したモデル。

表 9 時間関係認識に使用した素性の貢献度  
Table 9 Impacts of features.

	E2E		MAT		T2E		DCT	
	正答率	-Δ	正答率	-Δ	正答率	-Δ	正答率	-Δ
全素性使用 (Full')	67.13		50.00		70.34		80.72	
Full'-Base	64.50	-2.63*	49.87	-0.13	61.25	-9.09*	80.09	-0.63
Full'-Conjugation	-	-	-	-	-	-	80.02	-0.70
Full'-Value	-	-	-	-	70.28	-0.06	76.05	-4.67*
Full'-TypeClass	66.25	-0.88	49.87	-0.13	70.22	-0.12	80.34	-0.38
Full'-DepDistance	65.73	-1.40	-	-	67.17	-3.17*	-	-
Full'-Order	-	-	-	-	69.78	-0.56	-	-
Full'-ExistTime	66.87	-0.26	49.61	-0.39	70.03	-0.31	-	-
Full'-HeadTail	67.84	0.71	51.80	1.80	70.03	-0.31	80.75	0.03
Full'-FuncExpr	64.41	-2.72*	45.23	-4.77*	69.53	-0.81	79.39	-1.33*
Full'-Modality	68.01	0.88	50.39	0.39	70.03	-0.31	80.75	0.03
Full'-DepEvent	-	-	-	-	67.67	-2.67*	-	-
Full'-EventCount	66.96	-0.17	49.87	-0.13	-	-	-	-

\* Full' を比較対象とした 5-fold cross-validated paired t test による両側検定で、有意水準 0.05%以上の有意性を示したモデル。

た素性名単位で 1 つ素性を取り除いた際の正答率である。

表 9 の各素性を取り除いた際の正答率の差 (-Δ) を見ると、特に基本形 (Base)、係り受け距離 (DepDistance)、機能表現 (FuncExpr) の素性が正答率の向上に貢献していることが分かる。これらの素性には表 9 に示すとおり有意性が認められるタスクも存在した。基本形や機能表現の素性は、タスクに依存せず時間関係ラベルを推定するための有効な手がかりと分かった。また係り受け距離の素性は、同一文内の表現間の時間関係を推定する情報として有益であることが確認された。

一方で、解析対象の表現が文頭・文末に存在 (HeadTail)、拡張モダリティ (Modality) の素性は、一部のタスクで正答率に悪影響を与えていることが分かった。文頭・文末の素性による正答率の低下は、解析対象の表現間が連続して出現する E2E、出現位置が文末に限定されている MAT、構造的な素性が時間関係認識に影響しにくいと予想される DCT で見られた。このことから、コーパス上で文構造に関する制約が強いタスクでは、文頭・文末の素性は時間関係の推定に有用な情報となりえなかったと推測される。また、拡張モダリティは、E2E、MAT、DCT で悪影響を出

している。拡張モダリティ解析自体に誤りが生じている、もしくは現状の拡張モダリティのタグ体系では時間関係認識に有益な情報が得られていないことが原因として考えられる。これらの仮説を検証するためには、正しい拡張モダリティの値を手で与えたうえでの分析が必要である。

また我々が新しく提案した素性では、時間表現の係り先の事象表現の文脈情報をとらえる素性 (DepEvent) が有意性を示したことから、時間関係認識に効果的であることが分かった。しかし、大規模データから獲得した事象表現対の頻度情報の素性 (EventsCount) は、正答率の向上が見られたものの有意性が認められなかった。そこで獲得した事象表現対の頻度情報を人手で調査したところ、頻度上位 100 件のうち、約 6 割には時間的な順序関係の偏りを認めることができたが、パターン獲得の際に事象表現対の周辺文脈を省略したため、特定の時間的な偏りを断定できない事象対も多数存在した。また、頻度の高い事象表現対は「行う」などの汎用的な語や「終わる」などの表 7 の手がかり表現と共起しやすい語に偏っており、「催促する」などの Web 文書上で出現しにくい語に関しては少量しかパターンを獲得できなかった。よって、事象表現対のみでは



表 10 Mirza ら [29] との素性貢献度の比較

Table 10 Comparison of feature contributions with Mirza et al. [29].

	event-event (Mirza ら)		E2E + MAT (本研究)		event-timex (Mirza ら)		T2E + DCT (本研究)	
	正答率	直前との差	正答率	直前との差	正答率	直前との差	正答率	直前との差
majority	22.17		46.34		36.42		61.97	
string	31.07	8.90	56.47	10.13	58.27	21.85	68.55	6.58
+grammatical	36.15	5.08	56.74	0.27	61.30	3.03	69.52	0.97
+textual_context	39.44	3.29	-	-	61.71	0.41	70.19	0.67
+entity_attribute	40.47	1.03	59.09	2.35	64.60	2.89	75.26	5.07
+dependency_relation	42.63	2.16	59.35	0.26	65.22	0.62	76.48	1.22

表 11 Mirza ら [29] が使用した素性との対応

Table 11 Correspondence of features between Mirza et al. [29] and our work.

Mirza ら	本研究
string	Base
grammatical	POS, POS1, ConjForm
textual_context	Order
entity_attributes	Value, TypeClass, FuncExpr, Modality
dependency_relation	DepDistance, ExistTime, HeadTail

なく周辺文脈も考慮したうでの頻度抽出や、係り受けパターンに用いる表 7 の手がかり表現の増加など、事象間知識の獲得手法を改善していかなければならない。

### 5.2 日本語と英語間の素性の貢献度比較

表 10 は、本研究で作成した既存素性の貢献度と Mirza らの素性の貢献度分析 [29] を比較した結果である。ただし、本研究と Mirza らの間では使用した素性セットやタスク設定が異なるため、表 10 では本研究の設定を Mirza らの検証に合わせることで比較を図った。まず素性の差異については、Mirza らと本研究との間の素性の対応関係を表 11 のように定め、素性 string から素性を 1 つずつ追加した。次にタスク設定の差異に関しては、Mirza らの検証である event-event と event-timex が、それぞれ本稿の E2E と MAT を合わせたもの (E2E + MAT)、T2E と DCT を合わせたもの (T2E + DCT) と類似するため、それぞれのタスクの正答率の平均値を用いて Mirza らの検証と比較した。上述以外にも、時間関係ラベルの種類数が異なるなど多数の差異が存在するため、厳密な比較を行うことは不可能であるが、この検証はあくまで日本語と英語間の素性の貢献度の傾向差を観察するためのものであり、厳密な正答率比較のための実験ではない点に留意されたい。なお、表 10 の majority は、各タスク内で最も多く出現する時間関係ラベル (以降、最頻出ラベルと呼ぶ) の割合である。すなわち majority は、そのタスク内のすべての事例の推定結果を最頻出ラベルとして出力した際の正答率と見なすことができる。ただし日本語では、各タスクでの majority

の平均値を E2E + MAT および T2E + DCT の majority とした。

表 10 より、string と entity\_attribute は日本語と英語ともに正答率の上昇が見られたため、言語間共通で時間関係認識に有効な素性であると分かる。さらに T2E + DCT では、dependency\_relation の素性が有効に働いていることから英語との共通性が見られたが、E2E + MAT では dependency\_relation の素性で英語ほどの正答率向上が見られなかった。これは MAT は文間の表現間が解析対象であるため、dependency\_relation に該当する文構造の素性を導入できず、さらに文内の事象表現間を解析対象とした E2E で、効果的に働く文構造の素性が十分に存在しなかったためと考えられる。一方 T2E では、表 2 のとおり同一文内の時間表現と事象表現間が解析対象であり、解析対象の表現間に文構造に関する強い制約が存在しないため、dependency\_relation の素性が効果的に働いたと考えられる。

また grammatical は、英語では正答率が 3~5% 程度上昇しているが、日本語での正答率向上は乏しかった。英語では品詞体系として Penn Treebank のタグセットが用いられており、動詞の品詞解析の中で、現在形 (VB)、動詞の語尾に-ed が付いた過去形 (VBD)、動詞の語尾に-ing が付いた現在進行形 (VBG) といった事象の持つテンスやアスペクトなどの時間情報も解析される。これらは時間関係認識に有用な意味情報であると考えられるため、正答率の向上に大きく貢献したと推測される。一方日本語の品詞体系では、第 1 階層は「動詞」「名詞」「助動詞」、第 2 階層は「サ変接続」「自立」「非自立」「接尾」などの直感的に時間関係の推定に有用でないタグで構成されており、日本語では英語と比べて品詞の素性が効果的に働かなかったと考えられる。

英語では品詞解析から得ているテンス・アスペクトなどの時間情報は、日本語では事象に付随する「た」や「ている」といった機能表現が担う。そのため、日本語の時間関係認識では時間情報を持つ機能表現を正しくとらえることが重要であるといえる。実際に表 9 より、機能表現の素性が日本語の時間関係認識に高い効果をもたらしていることが確認できる。

表 12 先行研究との正答率の比較

Table 12 Accuracy comparison to previous work.

	E2E	MAT	T2E	DCT
majority	51.14	39.30	50.72	68.26
本研究	68.18	52.45	70.34	80.89
吉川ら [10]	*12(59.9)	50.0	55.7	75.6

このように、日本語と英語間の言語差に依存しない素性も存在する一方で、品詞解析の結果にテンス・アスペクトなどの時間情報を付与しにくい日本語特有の文法規則などの影響によって、英語とは異なる結果となった素性も存在した。よって、英語で採用されている素性を導入するのみではなく、本研究で採用した機能表現の素性のよう、日本語の特性に合わせた調整や日本語独自の素性の採用も吟味していかなければならない。

以降の実験では、5.1 節の表 8 と表 9 で悪影響を与えていた素性、すなわち E2E と MAT では「POS, POS1, ConjForm, HeadTail, Modality」、T2E では「POS, POS1, ConjForm」、DCT では「POS, POS1, HeadTail, Modality」を取り除いたモデルを使用した。

### 5.3 先行研究との比較

表 12 は、我々のモデルの正答率を、最頻出ラベルの割合 (majority)、日本語での先行研究にあたる吉川ら [10] のモデルと比較したものである。吉川らは時間関係ラベルを 6 種類にグループ化したうえで、5 分割交叉検定による評価を実施しており、本研究の実験設定とほぼ等価と見なせるため、正答率で直接比較した。しかし、吉川らは E2E を文内と文間の事例を区別せずに扱っているため、E2E では本研究と直接比較できない点に注意されたい。

タスク内のすべての事例の推定結果を最頻出ラベルとして出力する majority は、最も単純なベースラインの 1 つであるが、このベースラインと比較して我々のモデルは約 10~20% 高い正答率を達成した。英語と同様に機械学習による時間関係ラベルの分類は、日本語における時間関係認識でも十分に有効であることが分かった。

吉川らのモデルと比較すると、我々のモデルは直接比較可能なタスクにおいて吉川らより高い正答率を達成していることが分かる。特に、T2E では F1 値で 10% 以上と大きく上回っている。その主な理由として、我々のモデルでは、吉川らが採用していない解析対象の表現対の周辺に存在する基本形や時間表現の係り先の事象表現の周辺文脈を使用した素性を用いているためと考えられる。実際に表 9 で、Base や DepEvent の素性が強く正答率向上に寄与してい

\*12 表 12 の括弧内の値は、E2E の文内と文間の問題を合わせた問題での正答率である。仮に我々のモデルの MAT を E2E の文間の正答率と同等としたとき、E2E の文間の事例数が 724 であることから、我々のモデルでの E2E の文内と文間の正答率は 62.06 となった。

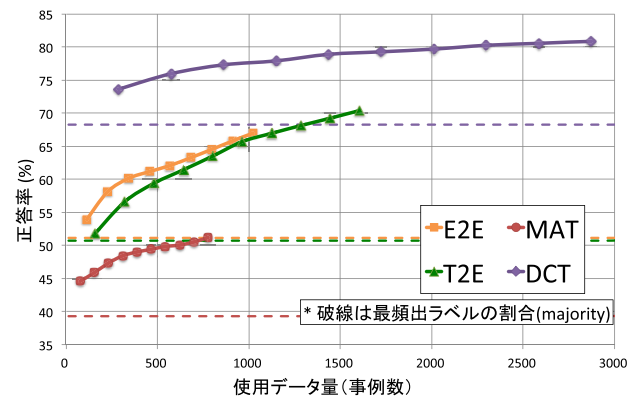


図 5 学習曲線

Fig. 5 Learning curves.

ることが確認できる。

### 5.4 学習曲線の分析

図 5 は、表 5 のデータ量を 10~100% の間で 10% ずつデータ量を増加させ、各タスクで 5 分割交叉検定を実施した際の学習曲線である。なお図 5 では、縦軸が正答率、横軸がデータ使用量、破線が最頻出ラベルの割合 (majority) を示している。

図 5 より、すべてのタスクで学習データ量増加にともなう正答率の飽和が発生していないことが確認できる。したがって、現状のコーパスでは学習データ量が不十分であり、仮により多くの学習事例が獲得できれば、さらなる正答率の向上が期待できる。

またタスク別に観察すると、特に T2E と E2E ではデータ量増加にともなう顕著な正答率の変化が見られることから、学習データ量の増加によって大幅な正答率の向上が期待できる。しかし、MAT や DCT ではデータ量増加にともなう正答率の上昇が小さく、既存の素性のみではすぐに限界に達することが推察される。実際に、MAT は文間の事象表現間が解析対象、DCT は文書作成日時と文書中の事象表現間が解析対象のタスクであり、同一文内の表現間が解析対象である E2E や T2E と比べると導入できた素性の種類が少なかった。よって、今後は MAT や DCT でも採用可能な素性の種類数を増加させていく必要がある。

一方で、いずれのタスクにおいても学習データ量が正答率の上昇に直結するものの、時間関係ラベルの付与は、一般に作業員間一致率を高く保つことが難しいことが知られており [4], [9]、本研究で利用した BCCWJ-TimeBank のような高品質の学習データの作成に高いコストが必要となる。よって今後の課題として、高品質の学習データを確保する手段についても吟味しなければならない。

### 5.5 各時間関係ラベルの推定精度

表 13 に各タスクの各時間関係ラベルの F1 値を示す。表 5 との対応から、どのタスクにおいても最頻出ラベル

表 13 時間関係ラベルごとの F1 値

Table 13 F1 scores of each temporal relation label.

	E2E	MAT	T2E	DCT
AFTER	49.68	48.37	71.19	88.61 <sup>1</sup>
BEFORE	79.09 <sup>1</sup>	60.6 <sup>1</sup>	61.85	69.11
B-or-O	*	*	16.84	*
O-or-A	*	*	12.38	*
OVERLAP	57.8	48.49	76.59 <sup>1</sup>	51.37
VAGUE	0	0	62.12	0

<sup>1</sup> 各タスクにおける最頻出ラベルを意味する.

\* 使用データ中に 0~2 しか存在しない (表 5 参照) た  
め, 5 分割交叉検定による評価が行えない.

の F1 値が高いことが確認できるが, これらの F1 値も実用を見据えると十分に高い値とはいえない. また, 事例数が 200 程度しか存在しない時間関係ラベルの F1 値は, 最頻出ラベルの F1 値と比較して 15~30%低いことも確認できる. 一方で, B-or-O や O-or-A の事例数が少量存在する T2E では, 微量ながら解析できているため, 他のタスクでも B-or-O と O-or-A のデータ量の増加でこれらの時間関係ラベルの推定が期待できる. これらの仮説を実証するため, 事例数の少ない時間関係ラベルのデータ量を増加させたうえでの検証が必要となる.

また時間関係ラベル VAGUE も, 事例数が少ないためにうまく学習ができていないと考えられるが, T2E では VAGUE も推定できていた. これは, たとえば「週 2 日」などの時間表現は, 時間軸上で複数の地点を指し示す表現であるが, ある事象表現との時間関係を考える際, 特定の 1 つの時間関係を決定することができない. このような特徴的な表層文字列から VAGUE かどうかを判断しやすいため, T2E では一部の VAGUE が解析できたと考えられる.

### 5.6 誤り事例分析

時間関係ラベルの解析精度に関する定量的評価のみではなく, 誤り事例に対する調査を実施した. 使用データを学習と推定に 4 対 1 で分け, 時間関係ラベルの推定の際に生じた誤り事例を調査の対象とし, 入力として与えられた解析対象の表現を含む文から正しい時間関係ラベルを認識するために, どのような情報をとらえる必要があるかを著者が人手で調査した.

調査の結果, 誤り事例を正しく解析するためにとらえる必要がある情報は, 表 14 に示す 6 つのカテゴリに分類された. ただし, 各事例は複数のカテゴリに属する場合がある. 以降では, 各カテゴリの詳細を説明するとともに, その代表的な事例を紹介する. なお, 事例の説明では時間関係認識の解析対象である表現対を下線で表現し, 正解の時間関係ラベルと我々のモデルが予測した時間関係ラベルを例文の末尾に記す. また, 現象の説明に影響を与えない範囲内で, 文書を簡略化した.

表 14 誤り事例の分類

Table 14 Categorization of errors.

	E2E	MAT	T2E	DCT
事象の知識・推論	21	16	8	7
事象の機能的意味	14	19	13	18
周辺文脈関連	10	15	45	35
文構造関連	16	3	21	9
入力からの判別困難	11	23	13	31
その他	8	7	9	15
誤り事例の総数	66	62	95	102

1 つ目の「事象の知識・推論」は, 正しい時間関係ラベルの推定に, 事象間のスクリプト的な知識 [1] や同義語・類義語関係などの 2 つの事象間の関係が記述された知識を利用し, 文書上で明示的に記述されていない情報を推論によって補うことが必要となる事例である.

(1) A 氏が自動車のドアに手を 挟んで 指を 骨折した ため, ... (正解: BEFORE, 予測: OVERLAP)

(1) の事例では, 事象表現のスクリプト的な知識を用いた推論によって, 正しい時間関係ラベルを導出できると考えられる. 「手を挟む」という事象により怪我が生じ, また「骨折する」にはその原因となる事象が存在するといった知識が存在するならば, この事例は「挟む」の後に「骨折する」という時間関係であると推論できる. 我々は, このような知識を「外部知識・大規模データの頻度情報」の素性でとらえようとしたが, 5.1.2 項の検証と同様に, 誤り事例の分析からも現状の手法では不十分と判断できる.

(1) は比較的単純な事例であるが, 誤り事例の中にはより複雑な推論が必要である事例や, 推論のために人間が常識的に使用している知識の活用が必要となる事例も見られた.

(2) A の像が市民の前で 倒された. 作業が難航し, 軍が車両で 手助け. (正解: AFTER, 予測: BEFORE)

(2) の事例では, 「(事象 X が) 難航する」の後「(事象 X を) 手助けする」といったスクリプト的な知識, 「難航」している事象は「まだ完了していない」といった事象の語彙的な知識, 「手助け」は困難な事象の補助をし, 補助した事象が完了に向かう傾向があるという常識的な知識などを組み合わせた推論によって, 「倒された」と「難航し」の間の正しい時間関係ラベルを効果的に判定できると考えられる. なお, この事例では「作業」が事象「倒す」を指し示しているが, 周辺文脈によっては「作業」が「倒れた結果生じる事象 (たとえば, 倒壊に巻き込まれた人命救出)」を指すことも許容できるため, 文脈に沿った正しい時間関係ラベルの判断には, 周辺文脈が必要となる複合的な事例である点に注意されたい\*13.

\*13 (2) の事例は, 表 14 における「入力からの判別困難」にも分類されている.



また、上位・下位語や類義語関係などの事象間知識が時間関係認識の手がかりになりうる事例も見られた。

- (3) 郊外のレストランでつくるのは、オリーブ油でからっと揚げたイワシ料理，…（正解：OVERLAP，予測：BEFORE）

(3)の事例では、「揚げる」は「つくる」の中で発生する事象の1つであることが分かる。このような場合、「つくる」が「揚げる」の上位語であるという事象間知識が、正しい時間関係ラベルの推定に有用な判断材料となる。2.2節で述べたが、英語ではD'Souzaらが既存の知識リソースを時間関係認識に利用しており、機械学習とルールベースの手法を組み合わせることで、その有効性を示している[28]。ただし、SVMによる機械学習手法のみの場合では、知識リソースを取り入れることで推定精度に悪影響を与えていたため、日本語に適用する際にも知識リソースの効果的な活用法を調査する必要がある。

これらの点から、知識の獲得手法の改善や既存の知識リソースの活用とともに、事象知識を効果的に使用方法も吟味していかなければならない。

2つ目の「事象の機能的意味」は、テンス・アスペクトや否定・仮定・モダリティなどの観点から事象表現に対する機能的意味の解析が必要な事例である。最も多く見られた事例は、事象表現のアスペクトの意味を表す「ている」が影響するものであった。

- (4) 患者にとっては不用意な告知となりかねないという課題も指摘している。（正解：AFTER，予測：OVERLAP）

機能表現「ている」は継続の意味を持つため、他の表現と時間的な重なりを持ち、OVERLAPが付与される傾向があり、実際に(4)の事例でも、誤ってOVERLAPと推定してしまった。しかし、「ている」には継続の意味に加え、結果状態や習慣などの意味で使用されることがあるため、表層からではなく「ている」の持つ意味を解析したうえで用いるべきである。またアスペクトと同様に、否定や仮定、モダリティ、未来・過去といったテンスなどの機能的意味も時間関係の推定と関連するため、これらに対する意味解析も考慮する必要がある。

3つ目の「周辺文脈関連」は、入力として与えられた解析対象の表現対のみではなく、文書内の他の表現に関する情報をとらえることが求められる事例である。その一例として、解析対象以外の時間関係を解析の手がかりとして用いることによって、正しい時間関係ラベルを推定できる事例が見られた。

- (5) 八日に出演したニュースの番組で、戦争前の二月に機密情報を開示したことについて述べた。（正解：AFTER，予測：OVERLAP）

本研究では文書内の時間関係ラベルをそれぞれ独立に推定したが、(5)のような事例では解析対象以外の時間関係が時間関係認識の重要な手がかりとなりうる。まず、文書中の時間表現を正規化した値(value)は、文書作成日時(2003-06-10)を考慮すると、「八日」が「2003-06-08」、「二月」が「2003-02」であり、これらから「八日」は「二月」より未来の日付であることが分かる。このとき、「開示する」が「二月」中に発生した出来事と断定できれば、機械的に「八日」の前に「開示する」が発生したことが計算できる。しかしこれらの計算では、「八日」と「二月」などの文書中の時間表現の同定とその正規化が前提となるため、時間関係認識の前処理として時間表現に対する時間情報処理が必要となる。また、「二月」と「開示する」などの他の時間関係認識の結果に依存するため、どの表現対から解析していくべきかという疑問点も生じるが、この事例からは、直接係り受け関係にある表現間の時間関係を優先的に推定していく方が有益であるようにとれる。時間表現の正規化ツールや解析対象以外の時間関係を素性として取り入れる、もしくはMarkov Logicなどの全体最適化手法を適用することで、複数の時間関係の相互作用を考慮することの有効性の確認が不可欠である。

また、文中に存在する明示的な時間的順序関係を示す表現をとらえることが、正しい時間関係ラベルの推定の手助けとなりうる事例も存在した。

- (6) … Xが倒された瞬間をテレビで見たAは…（正解：OVERLAP，予測：BEFORE）

(6)の事例では、「瞬間」という語によって、「倒される」と「見る」に同じ時間帯に発生していることが明示的に示されている。「瞬間」のほかにも、時間関係を判断するための手がかりとなりうる表現として「次、までに、から、以来、予定」なども確認された。我々のモデルでは、解析対象の表現対の周辺文脈をある程度とらえる素性を採用していたが、時間関係に関わる明示的な手がかりなどをとらえる素性は導入していなかった。より効果的な時間関係ラベルの推定には、どのような表現が明示的な手がかりとして時間関係に関連を持つかを吟味し、時間関係認識上でうまく扱うための設計が求められる。

4つ目の「文構造関連」は、文構造に関する情報が不足しているために誤りが生じており、さらなる文構造の解析によって正しい時間関係を推定できると考えられる事例である。このカテゴリの代表的な事例として、(7)のような時間表現の影響範囲に関する問題が存在する。

- (7) 4月末の外貨準備高は3626億ドルとなり、前月末に比べると11億ドル増と、3カ月ぶりにプラスに転じた。（正解：OVERLAP，予測：BEFORE）

「転じた」は「4月末」に発生した出来事であるが、これ

らの表現間に「前月末、3ヶ月」といった時間表現が存在しているため、簡単な文構造の情報のみでは「転じた」がどの時間表現と被覆しているかを決定するのは難しい。このような事例では、時間表現が文書中の他の表現に影響を与えるか、影響している場合は被覆している事象表現の特定が必要となる。しかしながら、時間表現が事象表現と被覆するかを解析するためには、正確な係り受け解析に加え、文書中の時間表現が係り受け木上でどこまで影響しているかをとらえなければならない。時間表現がどの事象表現を被覆するかを付与したデータを作成し、時間表現が被覆する範囲の同定が可能かを確かめ、その解析結果を時間関係認識へ適用することが、今後の展望としてあげられる。

また、(8)のような連体節の構造をとる事例においても、より高度な文構造の解析が必要になると考えられた。

- (8) X党がA氏と対立するB党首の首相就任を要求したため、交渉は決裂。(正解：OVERLAP, 予測：AFTER)

誤り事例の中では、連体節内の事象表現は主語や目的語などの意味役割として関連を持つ事象と時間的な重なりを持つ場合が多いことが確認できた。実際に、D'Souzaらは述語項構造解析の結果を時間関係認識の素性として採用し、その有効性を確認している[28]。しかしながら、本研究では連体節をとらえる素性や、連体節を構成する事象表現の時間情報を用いた素性、述語項構造解析を用いた素性を導入しなかったため、これらの情報をとらえることができなかった。よって、日本語においても連体節を構成する事象表現の時間的な意味機能の調査や、時間関係認識に対する効果の検証などが必要である。

5つ目の「入力からの判別困難」は、本研究で入力として扱った解析対象の表現対やそれらを含む文などの局所的な情報のみでは、正しい時間関係ラベルの推定が困難であった事例である。実際にこれらの事例の重みベクトルを観察すると、態を示す「られる」、特定の機能表現の存在や文頭・文末を判定する素性で偽と判定されたもの、読点や括弧などを根拠として占有率の高い時間関係ラベルを付与している事例が多数見られた。したがって、解析対象の表現対が含まれていない文や隣接分との談話関係なども考慮し、より文書的に広い文脈情報を取り入れる必要があると考えられる。

6つ目の「その他」は、上述までのカテゴリに分類されない事例であり、より複雑で応用的な意味解析などが必要となるものである。以降では、このカテゴリに分類した事例をいくつか紹介する。

- (9) 資格をもとに高校で単位を認定したり、大学入試や就職に活用することを求めている。(正解：VAGUE, 予測：BEFORE)

(9)の事例では、「認定する」と「活用する」は文書作成日時時点で実際に発生していない事象であり、さらにそれぞれが並列となっているため、これらの間には時間関係が存在しないが、誤ってBEFOREのラベルを付与してしまった。しかし、このような事例の場合、それぞれの事象表現が並列となっていることや、それらが同じ時間軸中で発生しない出来事であることを解析し、時間関係が与えられないことを認識する必要がある。

- (10) 本業のもうけを示す営業損益が三百十一億円の赤字に、純損益も五百五十八億円の赤字となった。(正解：VAGUE, 予測：BEFORE)

(10)の事例における「示す」は、一般的な事実や説明を表す表現であり、実際に文書上で発生した出来事とは異なるため「なる」との具体的な時間関係は存在しない。「示す」のような一般的な事実を述べている事象表現であることを認識する必要があるが、表層などの簡単に取得できる情報から判断できない。(9)や(10)の事例で正しい時間関係ラベルの推定に必要となる事象表現の情報解析は、いずれも自然言語処理における応用的な解析に位置付けられるため、これらの事象表現間に時間関係が存在しないと判定することは困難な問題であるといえる。

## 6. おわりに

本稿では、日本語における頑健な時間関係認識の解析手法の確立を目指し、ドメインに依存しない解析手法を対象に分析を行った。BCCWJ-TimeBankを用い、機械学習による時間関係認識を日本語に適用したところ、日本語でも英語で採用されている大半の素性が有効であった。ただし、英語では品詞情報を用いた素性が有効であったが、日本語の品詞体系では英語の品詞体系のタグセットで扱っているテンス・アスペクトなどの機能的意味をとらえられないため、品詞による素性は正答率の低下を招いた。一方で、日本語ではテンス・アスペクトなどを担う機能表現を素性に用いた実験により、その有効性が示されたことから日本語の時間関係認識では機能表現を正しくとらえることが重要であると分かった。また、我々が新規に提案した時間表現の係り先の事象表現の周辺文脈をとらえる素性は有意に解析精度を向上させたが、大規模データからの頻度情報による素性は頻度の獲得方法に改善が必要であることが分かった。さらに、学習曲線や時間関係ラベルごとの推定精度の分析から、現状のコーパスでは学習データ量が不足していることと、文間の事象表現間や文書作成日時と事象表現間が解析対象である場合は、既存の基本的な素性のみでは学習の際の素性が不十分であることを確認した。誤り事例に対して人手による調査を実施した結果、事象のスク립ト的な知識や上位・下位概念などの事象間知識の拡充とそれらの使用方法の吟味、テンス・アスペクトや否定・



仮定・モダリティなどの事象の文法的意味理解のための解析, 特定の時間表現が被覆する事象表現の範囲を同定するための文構造解析などが, さらなる時間関係認識の発展のために必要であることが分かった. 今後は本研究の成果をもとに, 日本語における時間関係認識のさらなる解析精度の向上や, 推定した時間関係を談話関係や共参照関係の解析などのより応用的なタスクで活用する枠組みの構築に取り組んでいきたい.

謝辞 本研究の一部は, CREST の助成を受けたものです.

#### 参考文献

- [1] Chambers, N., Wang, S. and Jurafsky, D.: Unsupervised Learning of Narrative Schemas and their Participants, *Proc. 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pp.602–610 (2009).
- [2] Inoue, N., Ovchinnikova, E., Inui, K. and Hobbs, J.R.: Coreference Resolution with ILP-based Weighted Abduction, *Proc. 24th International Conference on Computational Linguistics*, pp.1291–1308 (2012).
- [3] Pustejovsky, J., Day, D., Ferro, L., Gaizauskas, R., Hanks, P., Lazo, M., Sauri, R., See, A., Setzer, A. and Sundheim, B.: The TIMEBANK corpus, *Proc. Corpus Linguistics 2003*, pp.647–656 (2003).
- [4] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G. and Pustejovsky, J.: SemEval-2007 Task 15: TempEval Temporal Relation Identification, *Proc. 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp.75–80 (2007).
- [5] Verhagen, M., Sauri, R., Caselli, T. and Pustejovsky, J.: SemEval-2010 Task 13: TempEval-2, *Proc. 5th International Workshop on Semantic Evaluation*, pp.57–62 (2010).
- [6] UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M. and Pustejovsky, J.: SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations, *2nd Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proc. 7th International Workshop on Semantic Evaluation (SemEval 2013)*, pp.1–9 (2013).
- [7] 高崎 聡, 飯田 龍, 徳永健伸: ドメイン知識を利用したレシビテキストにおけるイベント間の時間関係の解析, 情報処理学会自然言語処理研究会予稿集, Vol.NL-211-1, pp.1–7 (2013).
- [8] Inoue, N., Iida, R., Inui, K. and Matsumoto, Y.: Resolving Direct and Indirect Anaphora for Japanese Definite Noun Phrases, *Journal of Natural Language Processing*, Vol.17, No.1, pp.221–246 (2010).
- [9] Asahara, M., Yasuda, S., Konishi, H., Imada, M. and Maekawa, K.: BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text, *Proc. 27th Pacific Asia Conference on Language, Information, and Computation*, pp.206–214 (2013).
- [10] 吉川克正, 浅原正幸, 飯田 龍: BCCWJ-TimeBank を対象とした時間的順序関係の推定, 言語処理学会第 20 回年次大会発表論文集, pp.1103–1106 (2014).
- [11] 稲田和明, 松林優一郎, 乾健太郎: 同一文内の表現対を対象とした日本語における時間関係認識, 言語処理学会第 20 回年次大会発表論文集, pp.348–351 (2014).
- [12] Pustejovsky, J., Casta, J., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A. and Katz, G.: TimeML: Robust specification of event and temporal expressions in text, *Proc. 5th International Workshop on Computational Semantics (IWCS-5)* (2003).
- [13] Grishman, R. and Sundheim, B.: Message Understanding Conference-6: A brief history, *COLING '96, Proc. 16th Conference on Computational Linguistics*, pp.466–471 (1996).
- [14] DARPA\_TIDES: The TERN evaluation plan; time expression recognition and normalization, *Working papers, TERN Evaluation Workshop* (2004).
- [15] Allen, J.: Maintaining knowledge about temporal intervals, *Comm. ACM*, pp.832–843 (1983).
- [16] 橋本泰一, 中村俊一: 拡張固有表現タグ付きコーパスの構築—白書, 書籍, Yahoo!知恵袋コアデータ, 言語処理学会第 16 回年次大会発表論文集, pp.916–919 (2010).
- [17] 工藤真由美: アスペクト・テンス体系とテキスト—現代日本語の時間表現, ひつじ書房 (1995).
- [18] 保田 祥, 小西 光, 浅原正幸, 今田水穂, 前川喜久雄: 『現代日本語書き言葉均衡コーパス』に対する時間表現・事象表現間の時間的順序関係アノテーション, 自然言語処理, Vol.20, No.5, pp.657–682 (2013).
- [19] 小西 光, 浅原正幸, 前川喜久雄: 『現代日本語書き言葉均衡コーパス』に対する時間情報アノテーション, 第 2 回コーパス日本語学ワークショップ (2012).
- [20] Mani, I., Verhagen, M., Wellner, B., Lee, C.M. and Pustejovsky, J.: Machine Learning of Temporal Relations, *Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp.753–760 (2006).
- [21] Chambers, N., Wang, S. and Jurafsky, D.: Classifying Temporal Relations Between Events, *Proc. 45th Annual Meeting of the Association for Computational Linguistics*, pp.173–176 (2007).
- [22] Cheng, Y., Asahara, M. and Matsumoto, Y.: NAIIST.Japan: Temporal Relation Identification Using Dependency Parsed Tree, *Proc. 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp.245–248 (2007).
- [23] Hagege, C. and Tannier, X.: XRCE-T: XIP temporal module for TempEval campaign, *Proc. 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp.492–495 (2007).
- [24] Llorens, H., Saquete, E. and Navarro, B.: TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2, *Proc. 5th International Workshop on Semantic Evaluation*, pp.284–291 (2010).
- [25] UzZaman, N. and Allen, J.F.: TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text, *Proc. 5th International Workshop on Semantic Evaluation*, pp.276–283 (2010).
- [26] Kolya, A.K., Ekbal, A. and Bandyopadhyay, S.: JU\_CSE\_TEMP: A First Step towards Evaluating Events, Time Expressions and Temporal Relations, *Proc. 5th International Workshop on Semantic Evaluation*, pp.345–350 (2010).
- [27] Ha, E.Y., Baikadi, A. and Lester, C.L.J.C.: NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology, *Proc. 5th International Workshop on Semantic Evaluation*, pp.341–344 (2010).
- [28] D’Souza, J. and Ng, V.: Classifying Temporal Relations with Rich Linguistic Knowledge, *Proc. 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.918–927 (2013).



- [29] Mirza, P. and Tonelli, S.: Classifying Temporal Relations with Simple Features, *Proc. 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp.308–317 (2014).
- [30] 渡邊陽太郎, 浅原正幸, 松本裕治: 述語語義と意味役割の結合学習のための構造予測モデル, *人工知能学会論文誌*, Vol.25, No.2, pp.252–261 (2010).
- [31] 江口 萌, 松吉 俊, 佐尾ちとせ, 乾健太郎, 松本裕治: モダリティ, 真偽情報, 価値情報を統合した拡張モダリティ解析, *言語処理学会第16回年次大会発表論文集*, pp.852–855 (2010).
- [32] Laokulrat, N., Miwa, M., Tsuruoka, Y. and Chikayama, T.: UTTime: Temporal Relation Classification using Deep Syntactic Features, *2nd Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proc. 7th International Workshop on Semantic Evaluation (SemEval 2013)*, pp.88–92 (2013).
- [33] 松吉 俊, 佐尾ちとせ, 乾健太郎, 松本裕治: 拡張モダリティタグ付与コーパスの設計と構築, *言語処理学会第17回年次大会発表論文集*, pp.147–150 (2011).
- [34] 水野淳太, 成田和弥, 乾健太郎, 大竹清敬, 鳥澤健太郎: 拡張モダリティ解析器の試作と課題分析, *ALAGIN NLP 若手の会合同シンポジウム* (2013).
- [35] 坂地泰紀, 増山 繁, 酒井浩之: 新聞記事中の文が因果関係を含むか否かの判定, *電子情報通信学会技術研究報告* (2010).
- [36] Okazaki, N.: *Classias: A collection of machine-learning algorithms for classification* (2009).



乾 健太郎 (正会員)

東北大学大学院情報科学研究科教授。専門は自然言語処理, 人工知能。1995年東京工業大学大学院情報理工学研究科博士課程修了。博士(工学)。同大学助手, 九州工業大学助教授, 奈良先端科学技術大学院大学助教授を経て, 2010年より現職。2014年度情報処理学会論文誌編集委員長, 同年度より同学会自然言語処理研究会主査。言語処理学会, 人工知能学会, ACL, AACL 各会員。



稲田 和明 (学生会員)

2013年東北大学工学部情報知能システム総合学科卒業。2015年同大学大学院情報科学研究科博士前期課程修了。同年より同大学大学院情報科学研究科博士後期課程に進学, 現在に至る。意味解析の研究に従事。



松林 優一郎 (正会員)

2010年東京大学大学院情報理工学系研究科・コンピュータ科学専攻博士課程修了。情報理工学博士。同年より国立情報学研究所・特任研究員を経て, 2012年より東北大学大学院情報科学研究科・特任助教, 現在に至る。意味解析の研究に従事。言語処理学会, 人工知能学会, ACL 各会員。