『現代日本語書き言葉均衡コーパス』に対する時間情報表現・ 事象表現間の時間的順序関係アノテーション

保田 祥[†]・小西 光[†]・浅原 正幸[†]・今田 水穂[†]・前川喜久雄[†]

時間情報抽出は大きく分けて時間情報表現抽出，時間情報正規化，時間的順序関係解析の三つのタスクに分類される．一つ目の時間情報表現抽出は，固有表現・数値表現抽出の部分問題として解かれてきた．二つ目の時間情報正規化は書き換え系により解かれることが多い．三つ目のタスクである時間的順序関係解析は，事象の時間軸上への対応付けと言い換えることができる．日本語においては時間的順序関係解析のための言語資源が整備されているとはいえず，アノテーション基準についても研究者で共有されているものはない．本論文では国際標準である ISO-TimeML を日本語に適応させた時間的順序関係アノテーション基準を示す．我々は『現代日本語書き言葉均衡コーパス』(BCCWJ) の新聞記事の部分集合に対して，動詞・形容詞事象表現に TimeML の〈EVENT〉相当タグを付与し，その事象の性質に基づき分類を行った．また，この事象表現と先行研究(小西，浅原，前川 2013)により付与されている時間情報表現との間の関係として，TimeML の〈TLINK〉相当タグを付与した．事実に基づき統制可能な時間情報正規化と異なり，事象構造の時間的順序関係の認識は言語受容者間で異なる傾向がある．このようなレベルのアノテーションにおいては唯一無二の正解データを作ることは無意味である．むしろ，言語受容者がいかに多様な判断を行うかを評価する被験者実験的なアノテーションが求められている．そこで，本研究では三人の作業によるアノテーションにおける時間的順序関係認識の齟齬の傾向を分析した．アノテーション結果から，時間軸上の相対的な順序関係については一致率が高い一方，時区間の境界については一致率が低いことがわかった．

キーワード：時間情報処理，事象意味論，コーパスアノテーション

Temporal Ordering Annotation on ‘the Balanced Corpus of Contemporary Written Japanese’

SACHI YASUDA[†], HIKARI KONISHI[†], MASAYUKI ASAHARA[†],
MIZUHO IMADA[†] and KIKUO MAEKAWA[†]

Temporal information extraction can be divided into the following tasks: temporal expression extraction, time normalization and temporal ordering relation resolution. The first task is a subtask of a named entity and numeral expression extraction. The second task is often performed by rewriting systems. The third task consists of event anchoring. This paper proposed a Japanese temporal ordering annotation scheme and

[†] 人間文化研究機構国立国語研究所, National Institutes for the Humanities, National Institute for Japanese Language and Linguistics

performed annotations by referring to ‘the *Balanced Corpus of Contemporary Written Japanese*’ (BCCWJ). We extracted verbal and adjective event expressions as <EVENT> in a subset of BCCWJ and annotated a temporal ordering relation <TLINK> on the pairs of the above event expressions and time expressions obtained from a previous study (Konishi et al. 2013). The recognition of temporal ordering by language recipients tends to disagree compared to the normalization of time expressions. We should not regard making unique gold annotation data as an objective in such a situation. If anything, we should evaluate the degree of inter-annotator discrepancy by subjects of experiments. Then, we analysed inter-annotator discrepancies by three annotators in temporal ordering annotation. The result showed that boundaries of time segments barely exhibit any agreement, whereas the annotation of temporal relative ordering tendency exhibits good agreement by the annotators.

Key Words: *Temporal Information Processing, Event Semantics, Corpus Annotation*

1 はじめに

情報抽出や文書要約の分野において情報の可視化を目的として、テキスト中出现する事象表現の表す事象が発生した時区間 (Time Interval) を時間軸 (Timeline) 上に写像することが行われている。このためには、テキスト中出现する時間情報表現の正規化 (時間軸への写像) のみならず、対象となる「文書作成日時と事象表現」や「時間情報表現と事象表現」、「二つの事象表現」間の時間的順序関係を付与することが必要になる。

英語においては哲学者・言語学者・人工知能研究者・言語処理研究者が協力して時間情報を含む言語資源の整備を進めている (Pustejovsky, Hanks, Saurí, See, Gaizauskas, Setzer, Sundheim, Ferro, Lazo, Mani, and Radev 2003b). 哲学者・言語学者は言語科学として (a) テキスト中の事象表現とその時間構造を形式的にどのように記述するかを探究することを研究目的とする。人工知能研究者・言語処理研究者は工学研究として (b) テキスト中の事象表現や時間的順序表現を同定し抽出する機械的なモデルの開発や評価を研究目的とする。前者にとって (b) は手段でしかなく、逆に後者にとって (a) は手段でしかない。しかしながら、共通の目標として時間情報の可視化¹を掲げ、前段落にあげたリサーチクエスションに対して、「アノテーション」と呼ばれる研究手法により共有言語資源を構築する試みが行われている。

一方、日本語においては時間情報を含む言語資源の整備は、人工知能研究者・言語処理研究者によるものが多く、研究目的も (b) の手段としてのものが多い。機械的なモデルの開発や評価を目的とすることが多く、計算機上に実現しやすい時間情報表現の切り出しや正規化レベルのアノテーションにとどまっている (IREX 実行委員会 1999; 小西 他 2013)。時間的順序関係の

¹ ここで「情報の可視化」とは、工学的な自動処理によるもののみならず、言語科学における形式意味論研究も含む。

アノテーションを行うためには、アノテーション対象となる事象構造の意味論的な形式的な記述の作業が必要となる。人工知能研究者・言語処理研究者にとっての手段とされる研究目的 (a) が重要になる。

時間情報のアノテーションについては、英語のアノテーション基準 TimeML (Pustejovsky, Castaño, Ingria, Saurí, Gaizauskas, Setzer, and Katz 2003a) を元に国際標準化作業が行われてきた。成果物の ISO-TimeML は策定時に多言語に対してアノテーションすることを想定し、各言語の研究者がそれぞれ適応²作業を実施してきた。

本研究では、研究目的として哲学者・言語学者の (a) の立場を取り、『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese; 以下 “BCCWJ”) (国立国語研究所 2011) の一部に対し時間情報表現と事象表現の時間的順序関係を付与するために、事象表現の切り出しと分類を行った。時間情報表現アノテーションの形式的な基準である国際標準 ISO-TimeML の日本語適応作業を MAMA サイクル (Pustejovsky and Stubbs 2012) (Model-Annotate-Model-Annotate サイクル、詳しくは 2.1 節で説明) を通して実施し、時間的順序関係付与に適した事象表現分類を行った。さらに、複数人の時区間の時間的順序関係の認識の差異を評価することを目的として、Allen の時区間論理 (Allen 1983) (詳しくは 3.3 節で説明) に基づいたテキストに出現する時間情報表現と事象表現の時間的順序関係のアノテーションを複数人で実施した。MAMA サイクルを最小にし被験者実験的な設定でアノテーションを行い、得られたデータの傾向を分析し、複数人の作業者間の心的空間における時間構造の差異を評価した。

意味論レベルのアノテーションにおいて、多くの研究が形式意味論的な記述を目標とする。生成された言語を直接何らかの記号的な意味表現に写像するための方法論を確立するためにアノテーションの MAMA サイクルを実施するが、唯一無二の意味表現に写像することを目的とするためにアノテーション一致率という指標を良くする方向に最適化するきらいがある。一方、認知意味論の考え方においては、生成された言語表現を受容する人間の認知活動という要素を考慮し、人間の空間認知能力やカテゴリー化などの認知能力を評価する目的で、被験者実験などの研究手法が用いられている。テキストを刺激として与え、意味表現を記述させる被験者実験も広義のアノテーションと呼ぶことができる。

本研究では人間の時間的順序関係の認知能力の差異の評価を目的として、教示である MAMA サイクルを必要十分レベルに極小化した、被験者実験としてのアノテーションを行う。結果、時区間の境界の一致が困難である一方、時区間の前後関係については 69.5% の一致率でアノテーションできることがわかった。

以下本論文の構成について述べる。2 節では関連研究について述べる。3 節では策定した基準

² ここで「適応」とは生物学における“種の環境に対応する形質の有無”の意味ではなく、工学における“対象の特性に対応する仕様やパラメータなどの変更”の意味である。

について述べる。4 節で BCCWJ にアノテーションした順序関係ラベルの分析を行い、結果を報告する。5 節で本論文のまとめを行う。

2 関連研究

2.1 コーパスアノテーション

一般に言語の生産過程の産物であるアノテーションなしのテキストコーパスからは、言語の受容過程について直接的に調査することは困難である。言語の受容過程の調査には、生産されたテキストを受容する過程を記号化する必要がある。テキストコーパスに対し作業者が内容を理解して記号を付与するアノテーションは、工学研究者のベンチマークデータ作成だけでなく、人の言語の受容過程を記録する一研究手法としても利用可能である。

コーパスアノテーション作業には二つの基準を決める必要がある。一つはアノテーションをどのような形式で表現するかという形式的な基準である。アノテーション対象が文字間なのか文字列範囲なのか、対象に対しシングルラベルを付与するのかマルチラベルを付与するのか、対象間の関係が推移的なのか対称的なのか、大局的な構造として木をなすのか有向非循環グラフをなすのかなどを決定し、抽象化する必要がある。抽象化された形式は、インラインで記述するのかスタンドオフで記述するのかなどを基準として定める。この形式的な基準は、研究者間の相互利用性を高めたり、構造学習器を実現するための必要な抽象表現の仕様を決定するために利用される。関係する研究者があらかじめ議論をして標準仕様をコミュニティ駆動で策定したり、最初に策定された類似のアノテーションの形式をそのまま事実上の標準にしたりなど、標準化機関以外による何らかの標準化が行われることが多い。

もう一つはコーパスに出現する言語表現をどのような記号に割り当てるかという値割り当てについての基準である。アノテーションにおいては、個々の事例についてどの形式に割り当てるのかという基準が必要であり、一般に言語テストなどを作業者にに行ってもらいその判断に基づき記号に写像する基準が策定される。しかし、アノテーション作業の当初から完全に健全な基準を作成することは困難であり、基準の策定とアノテーション作業を何度も繰り返しながら基準を更新する。

Pustejovsky (Pustejovsky and Stubbs 2012) は、基準の策定方法を含めたアノテーション作業に二種類のサイクルがあることを示している。一つは MAMA サイクルで図 1 の左のようなサイクル³である。もう一つは MATTER サイクル (Model-Annotate-Train-Test-Evaluate-Revise サイクル) で図 1 の右のようなサイクルである。工学研究のように構造学習器を作成することを

³ 図は “Model and Guideline”-“Annotate”-“Evaluate”-“Revise” からなり “MAER サイクル” と呼ぶべきであるが、引用元の表現 “MAMA サイクル” をそのまま本稿でも採用する。

(Pustejovsky and Stubbs 2012) p.28 (Figure 1-11) より

(Pustejovsky and Stubbs 2012) p.24 (Figure 1-10) より

図 1 MAMA サイクルと MATTER サイクル

目的とする場合には MATTER サイクルを用いることが多いが、MATTER サイクルで構造学習器が構成できないアノテーション初期においては MAMA サイクルを用いることが多い。言語研究で現象そのものを観察する場合においては MAMA サイクルのみで閉じてアノテーションを行う傾向がある。

このようなアノテーションの基準とサイクルを考えた場合に、アノテーション基準の妥当性はどのように評価されるべきだろうか。形式的な基準においては利用者系により評価されるべきであり、当該基準を利用するコミュニティの規模などにより定量的に評価され、相互利用における障害の有無などにより定性的に評価されるだろう。後者の値割り当てとしての基準においては、構造学習器の構成を目的として研究を実施するのであれば、未知事例を含めた構造学習器の性能により評価されるだろう。一方、言語研究を目的とする場合には、アノテーション作業を行う指針である基準の妥当性は、成果物のアノテーションそのものによって評価されるべきである。アノテーション単体としての評価は一致率などの定量的な指標を提示することが可能であるが、言語研究のためのアノテーションにおいては、必ずしも一致率などを目的関数として最適化を行っているわけではない。このようなアノテーション基準の妥当性を評価するためには、MAMA サイクルの外側の言語研究者によって評論として行われるべきである。近年、均衡コーパスが整備され、コミュニティ駆動によりアノテーション対象の標準化が行われてきた。各機関で様々なレベルの言語情報のアノテーションが進められている。このような状況を鑑みると、MAMA サイクルの外側の言語研究者による評論の代わりに、他のアノテーションとの重ね合わせによる齟齬検出結果から、アノテーションそのものの妥当性評価が検証される可能性がある。

2.2 コーパスアノテーション基準の標準化

コーパスアノテーションの基準について、形式的な基準については標準化機関などが共有すべき規格を提案している。例えば、国際標準化機構 (International Organization for Standardization:

表 1 TC 37/SC 4 の作業部会

TC 37/SC 4/WG 1	Basic descriptors and mechanisms for language resources 言語資源に関する情報を記述するための作業部会
TC 37/SC 4/WG 2	Semantic annotation アノテーションと表現方法を議論する作業部会
TC 37/SC 4/WG 3	Multilingual information representation 多言語対訳テキストに特化した作業部会
TC 37/SC 4/WG 4	Lexical resources 言語資源そのものに関する作業部会
TC 37/SC 4/WG 5	Workflow of language resource management 言語資源管理の作業手順を議論する作業部会
TC 37/SC 4/WG 6	Linguistic annotation 言語情報アノテーションを議論する作業部会

ISO) の標準化技術委員会 (Technical Committee) TC 37 は “Terminology and other language and content resources” と題し、言語資源に関するさまざまな標準化を提案している。そのなかに分科会 (Subcommittee) が五つ設定されているが、TC 37/SC 4 が言語資源管理 (Language resource management; LRM) に関する国際規格の規定を行っている。TC 37/SC 4 は作業部会を六つ (表 1) 設定しており、さまざまな形式・出自の一次言語データに対するアノテーションや XML に代表される汎用マークアップ言語に基づくアノテーションの表現形式についての仕様記述言語を設計している。例えば、公開されている規格として、語彙表の規格 Lexical Markup Framework (LMF: ISO-24613:2008)、素性構造表現 Feature Structure Representation (FSR: ISO-24610-1:2006)、単語分かち書き (ISO-24615-1:2010 が一般、ISO-24615-2:2011 が日中韓言語)、統語論アノテーション Syntactic Annotation Framework (SynAF: ISO-24615:2010) がある。意味論的アノテーション規格は作業部会 TC 37/SC 4/WG 2 を中心にさまざまな Semantic Annotation Framework (SemAF) が提案されている。時間情報表現関連については、英語で策定された TimeML (Pustejovsky et al. 2003a) をもとに TimeML 開発者と作業部会 TC 37/SC 4/WG 2 が連携をとりながら SemAF-Time (ISO-24617-1:2012) TimeML を提案した。次の 2.3 節では、時間情報表現関連のアノテーションの研究動向を示す。

2.3 時間情報表現に関する研究動向

時間情報表現は哲学・言語学・人工知能研究・言語処理など複数分野の研究者により研究されてきた。

以下では言語処理関連の代表的な研究を俯瞰する。テキスト中の時間情報表現を分析する研究は大きく分けて時間情報表現抽出、時間情報正規化、時間的順序関係解析の三つのタスクに分類される。一つ目の時間情報表現抽出は、固有表現・数値表現抽出の部分問題として解かれ

てきた. 二つ目の時間情報正規化は書き換え系により解かれることが多い. 三つ目のタスクである時間的順序関係解析は, 事象の時間軸上への対応付けと言い換えることができる.

表 2 に英語と日本語を対象とした時間情報表現に関連する研究を示す.

英語においては, 評価型国際会議 MUC-6 (Grishman and Sundheim 1996) の一タスク固有表現抽出の中に時間情報表現の抽出が含まれている. MUC-6 で定義されている時間情報表現タグ <TIMEX> は日付表現 (@type="DATE") と時刻表現 (@type="TIME") からなる. アノテーション対象は絶対的な日付・時刻を表す表現にのみ限定され, “last year” などといった相対的な日付・時刻表現は含まれていない. この MUC-6 のアノテーション基準 <TIMEX> に対し, Setzer は時間情報表現の正規化に関するアノテーション基準を提案している (Setzer 2001). 評価型国際会議 TERN (DARPA TIDES 2004) では, 時間情報表現検出に特化したタスクを設定している. TERN で定義された時間情報表現情報タグ <TIMEX2> は, 相対的な日付・時刻表現, 時間表現や頻度集合表現が検出対象として追加されている. 時間情報表現の正規化情報を記述する ISO-8601 形式を拡張した @value 属性などが設計され, こちらも自動解析対象となっている.

その後, Pustejovsky らによりアノテーション基準 TimeML (Pustejovsky et al. 2003a) が提案されている. その中では, TERN で用いられている <TIMEX2> を拡張した <TIMEX3> が提案され, さらに時間情報表現と事象表現の時間的順序関係に関連づけるための情報 <TLINK> が付加される. これらの情報は人手でアノテーションすることを目的に設計され, TimeBank (Pustejovsky et al. 2003b) や Aquaint TimeML Corpus などの人手によるタグつきコーパスの整備が行われた.

これらのコーパスに基づく時間情報表現の自動解析 (Boguraev and Ando 2005; Mani 2006) が試みられたが, タグの情報に不整合があったり, 付与されている時間的順序関係ラベルに偏りが

表 2 関連研究

英語の時間情報表現に関する関連研究		
MUC-6 (Grishman and Sundheim 1996) (Setzer 2001)	評価型会議 基準	時間情報表現の切り出しのみ 時間情報表現の切り出しと正規化
TERN <TIMEX2> タグ	評価型会議	時間情報表現の切り出しと正規化
TimeML <TIMEX3> タグ	基準	時間情報表現の切り出しと正規化
TimeML <TLINK> タグ	基準	事象間の時間的順序関係
TimeBank	コーパス	TimeML 基準によるタグつきコーパス
Aquaint TimeML Corpus	コーパス	TimeML 基準によるタグつきコーパス
(Boguraev and Ando 2005)	解析手法	時間情報表現-事象間の時間的順序関係解析
(Mani 2006)	解析手法	二事象間の時間的順序関係解析
TempEval, TempEval-2, TempEval-3	評価型会議	時間情報表現-事象間／二事象間の時間的順序関係解析
日本語の時間情報表現に関する関連研究		
IREX	評価型会議	時間情報表現の切り出しのみ
拡張固有表現体系 (Sekine, Sudo, and Nobata 2002)	基準	時間情報表現の切り出しのみ
(橋本, 中村 2010)	コーパス	時間情報表現の切り出しのみ
(小西 他 2013)	基準／コーパス	時間情報表現の切り出しと正規化
本研究	基準／コーパス	時間情報表現-事象間／二事象間の時間的順序関係

あったりなど扱いにくいものであった (Boguraev and Ando 2006). 2007 年に開かれた SemEval 2007 の一タスク TempEval (Verhagen, Gaizauskas, Schilder, Hepple, Kats, and Pustejovsky 2007) では、時間的順序関係のラベルを簡略化し、人手で見直したデータによる時間的順序関係同定のタスクが行われた. このタスクでは、時間情報表現に対する正規化情報 @value 属性などがデータにあらかじめ付与されており、事象表現の時間的順序関係同定に利用できる設定になっている.

時間情報表現の自動解析に関する研究は英語中心に行われていたが、やがて言語横断的な研究が進められ、前の 2.2 節に示したような国際標準化がすすめられた. その成果物として、アノテーション形式の共有可能な基準として ISO-TimeML が策定された. その作業と並行して、評価型会議 TempEval-2 (Verhagen, Saurí, Caselli, and Pustejovsky 2010) が実施され、英語だけでなく、イタリア語、スペイン語、中国語、韓国語に関しても同様なデータを利用したタスクが設定された. 2013 年に開かれる SemEval-2013 のサブタスク TempEval-3 (UzZaman, Llorens, Derczynski, Allen, Verhagen, and Pustejovsky 2013) では、データの規模を大きくした英語、スペイン語が対象となっている.

海外においては、哲学者・言語学者・人工知能研究者・言語処理研究者が共有可能な言語資源を作成するという大義のもと、分野横断的に研究が進められている. さらに多言語に拡張すべく言語横断的に研究が進められている. このような状況のもと個々の研究について境界を明確に示すことは難しい.

次に日本語の時間情報表現に関する研究を示す. 日本語において、時間情報表現抽出はアノテーションのみならず、評価型会議による解析手法の検討が行われている. IREX (IREX 実行委員会 1999) の一タスクとして、固有表現抽出タスクが設定された. IREX の時間情報では、日付・時刻表現を対象にし、相対的な表現が定義に含まれている. 関根らは拡張固有表現体系 (Sekine et al. 2002) を提案し、辞書/オントロジやコーパスの作成などを行っており、BCCWJ にも同じ体系の拡張固有表現タグが付与されている (橋本, 中村 2010). 時間情報表現正規化については、小西らが TimeML に基づく <TIMEX3> 相当のタグを BCCWJ の一部に付与し、時間情報表現の正規化を行っている (小西 他 2013). しかしながら、日本語の時間情報表現と事象表現をひもづける時間的順序関係に関する研究は、著者らが知る限りない.

最後に、時間的順序関係アノテーションの目的について言及する. 工学研究者は (1) 時間情報を解析する構造学習器の構成やベンチマークデータの整備を目的としている. 一方、言語研究者は、(2) 事象表現の時間構造を表現する形式意味論としての記述体系の精緻化を目的としている. これらに対し、本研究は (3) 受容者としてのアノテーション作業という要素を考慮し、認知意味論的な分析を目的とする. (3) の目的のために、被験者実験的な設定のアノテーションを実施する.

2.4 アノテーション対象としての BCCWJ

本節ではアノテーション対象である BCCWJ について述べる.

約 1 億語規模の書き言葉均衡コーパスである BCCWJ は 2006–2010 年に整備され, 2011 年に国立国語研究所 (以下「国語研」と略す) から一般公開された. サンプリングの手法から生産サブコーパス・図書館サブコーパス・特定目的サブコーパスの三つに大きく分かれる. 生産サブコーパスは 2001–2005 年に出版された書籍 (PB)・雑誌 (PM)・新聞 (PN) により構成され, 生産実態に基づいてランダムサンプリングされている. 図書館サブコーパスは 1986–2005 年に出版された書籍 (LB) により構成され, 流通実態に基づいてランダムサンプリングされている. 特定目的サブコーパスは図書館サブコーパスで十分に集まりにくい, 白書 (OW)・Yahoo!知恵袋 (OC)・Yahoo!ブログ (OY)・国会会議録 (OM) など様々なレジスタのテキストが収録されている.

BCCWJ にはコアデータと呼ばれる約 110 万語からなる部分集合が設定されている. コアデータには人手により国語研規程の短単位・長単位単語境界, UniDic 品詞体系に基づく形態論情報, 文節境界などが付与されている. コアデータは生産サブコーパスから書籍 (PB)・雑誌 (PM)・新聞 (PN) が, 特定目的サブコーパスから白書 (OW)・Yahoo!知恵袋 (OC)・Yahoo!ブログ (OY) が収録されている. 表 3 に各レジスタのサンプルについての統計を示す. このコアデータに対し, 国内の様々な研究機関により, 係り受け情報・述語項構造・節境界・モダリティ情報・フレームネット知識など重畳的にアノテーションが行われている. しかしながら, 100 万語規模のコアデータ全てに対してアノテーションを実施することは困難である. そこで, コアデータの各サンプルに対してアノテーションの優先順位をつけ, 約 5–6 万短単位ごとの部分集合 (表 3・2 列目) を規定している. アノテーションに従事する研究者は, それぞれの目的や能力に応じ, この優先順位に従ってアノテーションを実施する. これにより, 優先順位の高いサンプルについてはより多種の言語情報アノテーションが行われることになる.

各サンプルには書誌情報として様々なメタデータが付与されているが, 本研究に重要なメタ

表 3 BCCWJ コアデータと部分集合

レジスタ	(略記号)	部分集合	サンプル数	短単位形態素数
白書	OW	A, B, C, D	62	197,011
書籍	PB	A, B, C, D	83	204,050
新聞	PN	A, B, C, D, E	340	308,504
Yahoo!知恵袋	OC	A, B	938	93,932
雑誌	PM	A, B, C, D	86	202,268
Yahoo!ブログ	OY	A, B	471	92,746
合計			—	1,098,511

データとして文書作成日時相当の情報がある。コアデータに収録されている 6 種類のレジスタのうち、新聞 (PN) データのみが日単位の文書作成日時の情報が収録されており、他のレジスタは年単位の文書作成日時の情報にとどまっている。

本研究では新聞 (PN) データの部分集合 A (54 ファイル⁴, 2,541 文, 56,518 短単位) を対象にアノテーションを行う。アノテーション作業対象を上記範囲に限定した理由は、BCCWJ のコアデータにおいて新聞データのみが文書作成日時を日単位まで保持していること、生産実態に基づいて適切にサンプリングされており通常の報道記事のみならずレシピやコラムが含まれていること、作業者が一人月でアノテーションを終えることが可能な分量であることなどがある。

3 アノテーション基準

3.1 アノテーション作業の概要

アノテーション作業対象は BCCWJ コアデータ新聞データ 54 ファイル (部分集合 A) とする。小西らの時間情報表現の正規化作業により、時間情報表現は〈TIMEX3〉タグにより切り出され、時間情報の正規化情報が与えられている (小西 他 2013)。

アノテーション作業は、最初に事象表現の境界を認定し〈EVENT〉タグを付与し、〈EVENT〉の属性として事象表現の分類を表す @class 属性を付与する。@class 属性付与の際には時間軸上に事象のインスタンスが認定できるか否かを判断し、判断できる場合には〈EVENT〉に対して〈MAKEINSTANCE〉タグをスタンドオフ形式で新たに付与する。次に限定された事象のインスタンス間 (「文書作成日時と事象表現」, 「時間情報表現と事象表現」, 「二つの事象表現」) に対して、時間的順序関係を付与する。以下では、それぞれの作業の基準について示す。

3.2 事象表現の認定とクラス分類

時間的順序関係のアノテーションを行うために、アノテーション対象である動詞・形容詞・形状詞が事象表現か否か、事象表現が時間軸上の特定の範囲で生起したものか否かの判断が必要となる。また事象構造が動作なのか状態なのかといった識別が必要になる。また、事象表現間の時間的順序関係を規定するにあたっては、ある事象が他の事象の項になりうるのか、その場合にどのような事象構造を持つのかを分類する必要がある。

国語研規程による長単位の動詞・形容詞・形状詞 4,953 表現に対して〈EVENT〉タグを付与する。事象表現として切り出す際に国語研長単位が適さない場合には切り出し範囲を大きくする方向で修正を行う。本研究は時間情報表現と事象のインスタンス間の時間的順序関係を付与するため、TimeML のアノテーションの形式的な基準に基づいて、実世界もしくは架空世界

⁴ BCCWJ において 1 ファイル中に複数の記事が収録されているために記事数ではない。

の時間軸上の具体的な特定の範囲で生じたインスタンスが認められるか否かの判別を行い、インスタンスが認められたものについては、〈EVENT〉タグの `@class` 属性にその事象表現の特性を付与し、〈MAKEINSTANCE〉タグを付与する。インスタンスが認められないものについては、〈MAKEINSTANCE〉タグを付与しない。

時間的順序関係が確認できる事象構造には〈MAKEINSTANCE〉タグを付与したうえで、`@class` 属性を付与する。`@class` 属性は、OCCURRENCE, REPORTING, PERCEPTION, ASPECTUAL, I_ACTION, I_STATE, STATE の 7 種類と作業者がインスタンスが認められないと判断した事象表現・静態表現に付与する NULL, NONE の 2 種類に分類される。

OCCURRENCE 項に事象を取らない事象表現一般

REPORTING 項に事象を取る表現活動動詞に相当する事象表現

PERCEPTION 項に事象を取る認識・知覚動詞に相当する事象表現

ASPECTUAL 項に事象を取るアスペクトを表出する事象表現

I_ACTION 項に事象を取る遂行動詞に相当する事象表現

I_STATE 項に事象を取る思考・感情動詞に相当する事象表現

STATE 静態動詞, 形容詞

NULL, NONE 時間軸上インスタンスが認められない事象表現

一般の事象表現は OCCURRENCE にあたる。静態動詞は STATE に分類されるため、STATE にしないもので事物 (Thing) を項とする事象表現はすべて OCCURRENCE とする。残りの 5 種類は、事物ではなく事象 (Event) を項として導入する事象にのみ用いる。なお、アノテーション対象としての事象は動詞・形容詞・形状詞に限定するが、項として事物か事象かを判断する際には事象名詞も考慮する。

この事象表現のインスタンスの認定とクラス分類は、作業者二人と監督者一人と助言者一人で議論しながら作業を行った。クラス分類を含めて 75-80% の一致率がコンスタントに得られるまで作業者二人が同一ファイルを作業し、基準が固まった時点で分担して作業を行った。基準の策定にあたっては日本語学・言語学の文献 (工藤 1995, 2004; 中村 2001) にある事象表現の分類を参考にした。

以下にそれぞれの例を挙げる。

OCCURRENCE：事象表現一般

何かが起こった、変化した、発生したなどの一般的な事象構造は、OCCURRENCE とする。すなわち、事象ではなく事物を項とし、静態動詞ではない場合は、すべて OCCURRENCE とする。無意志的 (状態・位置) 変化動詞や非意志的 (現象一般) 動詞もこれに含まれる。また、過程 (Process) を示す動詞 (例：「住む」) も、OCCURRENCE とみなすこととする。

〈EVENT〉@OCCURRENCE の例

湿地や干潟、河原などが埋め立てで〈EVENT〉減った〈/EVENT〉東京湾。
裸地を好むコアジサシに〈EVENT〉嫌われた〈/EVENT〉か、巣は一つだけ。
ニュース写真として〈EVENT〉掲載させていただく〈/EVENT〉ことがあります。
経常利益は数億円単位の黒字に〈EVENT〉なる〈/EVENT〉。
メニューに〈EVENT〉挑戦した〈/EVENT〉。

REPORTING：表現活動動詞

表現活動動詞が、事象に関する発言や告知などをはじめ、概ね「～と」を用いた引用を行う場合などで、REPORTING に分類する。なお、「～を」が用いられている場合は項が事物であるため、OCCURRENCE となる。表現活動動詞には、言う・報告する・告げる・説明する・陳述する・指摘する・伝えるなどが含まれる。

〈EVENT〉@REPORTING の例（太字が注目している項）

大学院でのこうした取り組みは初めてと〈EVENT〉いう〈/EVENT〉。
～どうかと〈EVENT〉提言する〈/EVENT〉。

PERCEPTION：認識・知覚動詞

認識動詞や知覚動詞で、主に事象に関する物理的な知覚が、節や句の「～の」などによる体言化によって導入される場合などは、PERCEPTION に分類する。但し、項が事物であるときは、OCCURRENCE とする（例「ホスピスという言葉を初めて聞いた」）。見る・観察する・見かける・眺める・聞く・聴く・耳にする・睨む・探る・感じるなどが含まれる。

〈EVENT〉@PERCEPTION の例（太字が注目している項）

母親が炊飯器でおでんを作ったのを〈EVENT〉見て〈/EVENT〉。

なお、新聞データにおいては、文脈により物理的な知覚を導入しない場合が多く、出現が少ない。

〈EVENT〉@PERCEPTION としない例（太字が注目している項）

[個人名]に[内容]について〈EVENT〉聞いた〈/EVENT〉。（インタビューであるため、OCCURRENCE）
A を B と〈EVENT〉見る〈/EVENT〉。（判断であるため、OCCURRENCE や I.STATE）

ASPECTUAL：アスペクト動詞

事象のアスペクト（相）を示す動詞が、事象を導入している場合はこれにあたる。明示的に記述されている場合に限定する。そのため、接頭辞などの造語成分（例：「再」＋動詞による「再団結する」「再開発する」など、「終」「開」による「終演する・開幕する」など）を含む動詞については、ASPECTUAL に含めない。

アスペクトを明示的に表す動詞は、以下のようなものがある。

- (1) Initiation：始める・始まる
- (2) Reinitiation：再開する
- (3) Termination：終える・止める・終わる・中止する・停止する・あきらめる
- (4) Culmination：やり終える・完成させる
- (5) Continuation：続ける・続行する・持続する・維持する・やり通す・保つ

〈EVENT〉@ASPECTUAL の例（太字が注目している項）

トーナメントは、日本時間 10 日夜に第 1 日が〈EVENT〉始まる 〈/EVENT〉。

〔個人名〕が勝てば **3 連覇**に〈EVENT〉続く 〈/EVENT〉偉業達成。

二年目も引き続き**好調**を〈EVENT〉維持したい〈/EVENT〉。

とろ火状態を〈EVENT〉保つ 〈/EVENT〉。

I_ACTION (Intensional Action): 内包的な動作

明示された事象の導入を行う（項とする）遂行動詞は I_ACTION と分類する。遂行しない場合は後述する I_STATE として区別を行う。また、イベントが助詞によって分割されている場合の後半部（例：「連絡をとる」「明らかにする」など）は I_ACTION と考える。次の I_STATE との差別として、挑む・予防する・遅らせる・依頼する・要求する・説得する・約束する・決定する・提案するなど、遂行性のある動詞がこれにあたる。また、同様に、REPORTING との差別として、宣言する・主張する・申し出る・断定するなど、PERCEPTION との差別として、調査する・精査するなどが I_ACTION にあたる。

なお、Intentional（意図的）とは異なることに注意されたい。

〈EVENT〉@I_ACTION の例（太字が注目している項）

女性が受け入れられるべきかと〈EVENT〉問われれ 〈/EVENT〉ば、イエスだ。

再建を国際社会全体で〈EVENT〉取り組む 〈/EVENT〉契機。

支払えないケースが〈EVENT〉出ている 〈/EVENT〉。

〔個人名〕は速い**転がり**を〈EVENT〉確かめていた 〈/EVENT〉。

I_STATE (Intensional States): 内包的な静態動詞

事象を導入する（項とする）が、事象を遂行しない動詞は I_STATE とする。代替・候補が言及されるなどの状態の導入が主となる。主に思考動詞や感情動詞がこれにあたり、信じる・思う・望む・欲する・期待する・計画するなどの思考動詞のほか、恐れる・心配する・悩むなどの感情動詞、また、遂行のない動詞として、求める・～しようとする・～したがるなど、～できる・～できないなども含まれる。

〈EVENT〉@I.STATE の例 (太字が注目している項)

連覇を〈EVENT〉狙う</EVENT>. 生活が〈EVENT〉できる </EVENT>.
未現像でも〈EVENT〉構いません</EVENT>.
よく見てくれたと〈EVENT〉感謝する </EVENT>. (遂行性がないため, I.ACTION ではない)

STATE : 静態動詞, 形容詞, 形状詞

時間的順序関係と直接かかわらない場合, 文書作成時間に従属しない場合には, 〈EVENT〉タグをつけないが, 以下の種類の静態動詞 (工藤 1995) と形容詞について, 時間と関わる場合に限り, 〈EVENT〉@STATE とする.

- (1) 存在動詞: ある・いる・存在する・点在する
- (2) 空間的配置動詞: そびえている・面している・隣接している
- (3) 関係動詞: 値する・あたる・あてはまる・相当する・意味する・示す・適する
- (4) 特性動詞: 甘すぎる・大きすぎる・泳げる・話せる・似合う

〈EVENT〉@STATE の例

マネジャーに就任する意向が〈EVENT〉ない</EVENT> ことを明らかにした. (存在)
東京湾岸でも生活〈EVENT〉できる </EVENT> 環境さえあれば. (特性動詞. この場合, 「生活ができる」であれば I.STATE とする)
彼女のようにモノをはっきり〈EVENT〉言える</EVENT> ことがこれからは大切だ. (特性動詞)
おいしく〈EVENT〉食べられます</EVENT>. (特性動詞)

NULL, NONE : 時間軸上インスタンスが認定できない事象表現・静態表現

〈MAKEINSTANCE〉を付与しない事象表現・静態表現. 〈MAKEINSTANCE〉タグを付与するか否かの判断基準として, 文書作成日時もしくは他の事象表現との時間的順序関係が定義できるかどうかを重要視する. 何らかの変化を含む事象表現ではなく, 恒常的あるいは一般的なことをいっていると考えられうる事象表現においては, 時間的順序関係のアノテーションは不可能であるため, 〈MAKEINSTANCE〉タグは付与しない.

〈MAKEINSTANCE〉タグを付与しない例 (太字が付与しない表現)

クラブの運営について 1 票を持っているわけではない.
国際会議 57 件を含め 2,111 件. 火を使わない調理法.

連体修飾節中の動詞が, 一般的と判断される場合 〈MAKEINSTANCE〉タグを付与しない.

〈MAKEINSTANCE〉タグを付与しない例: 連体修飾 (太字が付与しない表現)

旅の安全を守る道祖神. オリーブ畑に囲まれたレストラン.

副詞的用法や慣用的な場合も時間的順序関係が付けがたいため, 〈MAKEINSTANCE〉タグを付与しない.

〈MAKEINSTANCE〉タグを付与しない例：慣用表現（太字が付与しない表現）

やむを得ない、相次いで出している。
なりふり構わぬ販売攻勢。

文脈によっては、「ある」「なる」「する」などの動詞も、一般的なことを述べているため時間的順序関係が付けがたい場合がある。この場合〈EVENT〉タグを付与しない。

〈MAKEINSTANCE〉タグを付与しない例：「ある」など（太字が付与しない表現）

～のためこの名がある。～が基本となる。
～を原則とする。

以下、NULL と NONE のラベルの違いについて述べる。NULL のラベルは本節の作業を行った作業二人により付与したものである。NONE のラベルは次節の時間的順序関係認定時に、三人の作業者が時間軸上にインスタンスを認定することができなかったものについて修正付与する。

事象構造そのもののアノテーションは意味論レベルの情報付与に相当し、言語学的な知見から様々な記号化手法が考えられる。本研究は時間情報表現・事象表現間の時間的順序関係の可視化を目的としており、そのアノテーション形式の標準化である ISO-TimeML の枠組の範囲内で、値割り当てとしてのアノテーション基準を定めた。本節のアノテーション作業にあたっては次の 3.3 節で行う被験者実験的な時間的順序関係アノテーションの基底となる情報のために MAMA サイクルに基づき厳密な統制を行った。

3.3 時間的順序関係の認定

本研究は時間構造に対する複数人の認識の差異を評価するために、被験者実験的に時間的順序関係アノテーションを実施する。時間的順序関係について、事象が表現する時間構造が長さ 0 以上の時区間である（時点は長さ 0 の時区間として扱う）と仮定をおく。このことにより個人が認識する事象表現の時間構造を、人工知能分野でよく研究されている Allen の時区間論理 (Allen 1983) として表現することができる。アノテーション作業者は、時間軸上に二つの時区間をプロットすることで描画的に事象の時区間を表現することができる。直感的であるために短時間の教示でアノテーションが可能になる。

具体的には、先行研究で付与されている〈TIMEX3〉タグ範囲の時間情報表現と〈MAKEINSTANCE〉タグにより認定した事象表現のインスタンスに対して、〈TLINK〉相当の時間的順序関係を認定する。表 4 に示す Allen の二次の範囲代数に基づくラベル（13 種類）を付与する。採用するラベル集合は、標準化されているアノテーション形式であるため、他の研究者が多言語で言語横断的に分析する際にも有効だと考える。

なお、二つの事象表現が during/equal/contains の三つの時間的順序関係にある場合、部分事象の関係か全く同一の事象の関係でありうる。そのような場合には表 5 の三つのラベルを付与

表 4 Allen の範囲代数に基づく時間的順序関係ラベル

ラベル	意味
after	時間情報・事象表現 A が事象表現 B より後に起こる
met-by	時間情報・事象表現 A が事象表現 B の直後に起こる
overlapped-by	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より後、A の終了点は B の終了点より後である
finishes	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より後、A の終了点と B の終了点は同時である
during	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より後、A の終了点は B の終了点より前である
started-by	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点と B の開始点は同時、A の終了点は B の終了点より後である
equal	時間情報・事象表現 A と事象表現 B の時間的な重なりが完全に一致する
starts	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点と B の開始点は同時、A の終了点は B の終了点より前である
contains	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より前、A の終了点は B の終了点より後である
finished-by	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より前、A の終了点と B の終了点は同時である
overlaps	時間情報・事象表現 A と事象表現 B の間に時間的な重なりがあるが、A の開始点は B の開始点より前、A の終了点は B の終了点より前である
meets	時間情報・事象表現 A が事象表現 B の直前に起こる
before	時間情報・事象表現 A が事象表現 B より前に起こる

表 5 事象-部分事象間関係を表現するラベル

ラベル	時間的順序	意味
is.included	(during 相当)	事象表現 A は事象表現 B の一部（部分事象：subevent）である （時間的な重なりがあり、A の開始点は B の開始点より後、A の終了点は B の終了点より前である） 例えば、「卵を割る」は「オムライスを作る」と is.included の関係にある。
identity	(equal 相当)	事象表現 A と事象表現 B は全く同じ事象を示す（言い換え） 例えば、「オムライスを作る」と「オムライスを料理する」は identity の関係にある。
includes	(contains 相当)	事象表現 B は事象表現 A の一部（部分事象：subevent）である （時間的な重なりがあり、A の開始点は B の開始点より前、A の終了点は B の終了点より後である） 例えば、「オムライスを作る」は「卵を割る」と includes の関係にある。

する⁵。計 13+3 種類のラベルをまとめると図 2 のようになる。このほかにテキストの情報だけでは全く時間的順序関係がわからない場合に付与するラベルとして ‘vague’ を利用する。この作業は TimeML の〈TLINK〉付与の作業に相当し、我々もタグ名として〈TLINK〉を用いる。

これらの計 13+3+1 種類のラベルを〈TIMEX3〉タグと〈MAKEINSTANCE〉タグの間、もしくは、

⁵ 厳密には finishes/started-by/starts/finished-by の四つの時間的順序関係の場合も事象-部分事象関係になることがあるが、これらの関係の頻度が少なく相当する事例が見つからなかったことと TimeML には規定されていないことから、我々も相当するタグを新たに規定しない。

図 2 〈TLINK〉時間的順序関係ラベル一覧

二つの〈MAKEINSTANCE〉タグ間に付与する。本作業で用いる〈MAKEINSTANCE〉タグは前節の作業を精査した 3,839 件を固定して用いる。前節の作業を行った二人とは異なる、三人の作業者が時間的順序関係を行う。文書中の〈MAKEINSTANCE〉タグの対の数は文書中の〈MAKEINSTANCE〉タグの数の組み合わせに相当し、人手で全ての対を検証することは困難である。

英語の TimeBank では、全ての対で作業者が認定できるものという曖昧な基準で一致率が 55% と報告されている。本研究では、TempEval などの評価型ワークショップで採用されている「文書作成日時と事象表現の順序関係（“DCT” と呼ぶ）」「同一文内の時間情報表現と事象表現間順序関係（“T2E” と呼ぶ）」「隣接事象表現間順序関係（“E2E” と呼ぶ）」「隣接文の末尾の事象表現間順序関係（“MATRIX” と呼ぶ）」の 4 種類の表現対についてのみ付与する。

英語の TimeBank は、どの表現対に関係を付与するかというのは作業者にゆだねられている。一方、本研究では 4 種類の表現対について必ず何らかの関係を付与することとし、現実世界の事象と仮想世界の事象間、もしくは、二仮想世界の事象間などの場合で時間的順序関係が規定できない場合に ‘vague’ を付与することとしている。本作業の基準では 4 種類の表現対のうち “DCT”, “E2E”, “MATRIX” の 3 種類について複数の連結可能な単純道をグラフ上確保しており、基本的にアノテーションは連結グラフを構成する。このグラフ中 ‘vague’ の関係が切断辺となる場合、分離された部分グラフは二つの異なる可能世界（実世界-架空世界、異なる二架空世

界)を明示的に表現する.

なお, アノテーション作業に際し, 以下の点に注意した.

- 時間は基本的に区間としてアノテーションを行う. 1秒でも区間とする.
- 事象は瞬間動詞については点(長さ0の区間)とし, それ以外の表現は区間とする.
- 状態動詞などで開始点・終了点がわかりにくいものは, 前工程の〈EVENT〉タグの認定時で排除されているべきだが, わかりにくい場合には作業者の理解にゆだねる.

4 アノテーション情報の分析

4.1 事象表現の認定とクラス分類

時間的順序関係を行う前に時間情報表現と事象表現の範囲を切り出す必要がある. 時間情報表現の切り出しについては先行研究(小西 他 2013)によりなされており, 今回対象のBCCWJコアデータ新聞データ 54 ファイル上の分布は表6のようになっている. 事象表現の認定とクラス分類の分布は表7に示す.

表 6 時間情報表現の分布

時間情報表現分類	(〈TIMEX3〉@type)	件数
文書作成日時	(DATE)	54
日付表現	(DATE)	727
時刻表現	(TIME)	107
時間表現	(DURATION)	291
頻度・集合表現	(SET)	19
合計		1,198

表 7 事象表現の分布

	〈EVENT〉@class	件数
インスタンスが認められない事象表現 (2.2 節作業)	NULL	1,114
インスタンスが認められない事象表現 (2.3 節作業)	NONE	15
項に事象を取らない事象表現	OCCURRENCE	2,352
項に事象を取る事象表現	(5 種類全て)	(1,291)
	REPORTING	126
	PERCEPTION	27
	ASPECTUAL	63
	I_ACTION	880
	I_STATE	195
静態表現で事象表現として認めるもの	STATE	181
合計		3,839

4.2 時間的順序関係の認定

作業員三人により時間的順序関係認定作業を開始した。計 13+3+1 種類のラベルを「文書作成日時と事象表現の順序関係 (“DCT”）」「同一文内の時間情報表現と事象表現間順序関係 (“T2E”）」「隣接事象表現間順序関係 (“E2E”）」「隣接文の末尾の事象表現間順序関係 (“MATRIX”）」の 4 種類の表現対に対して付与した。

以下、作業員三人分の作業結果を示し、考察する。表 8 が 13+3+1 種類のラベルと 4 種類の表現対ごとに集計したものである。∩ で結ばれた三つの数字は、三人の作業員が何件その関係を認定したかを示す。右 “=” 以下の数字は其中で三人が一致した件数を示す。まず、一致したラベルの件数として、始点・終点の一致を必要としない ‘after’, ‘during’, ‘contains’, ‘before’ の頻度が多かった。始点・終点のいずれかの一致を必要とするラベルのうち、もっとも一致件数が多いものは時間軸上の完全の一致を示す ‘equal’ であった。また ‘vague’ についても複数の作業員が認定し、314 件一致しているところから、文脈を用いても時間的順序関係が推定できないものが少なからずあることがわかる。

表 9 に 4 種類の表現対ごとの一致率を集計したものを示す。一致率の評価基準として、「ラベル 13+3+1 種類を区別するもの (ラベル 13+3+1)」「部分集合であるか否かを区別せず、ラベル 13+1 種類を区別するもの (ラベル 13+1)」「TempEval で用いられているラベル 5+1 種類 (‘BEFORE’, ‘BEFORE-OR-OVERLAP’, ‘OVERLAP’, ‘OVERLAP-OR-AFTER’, ‘AFTER’, ‘VAGUE’⁶) に縮退するもの (ラベル 5+1)」の 3 種類を用いる。まず、もっとも厳しい一致率

表 8 〈TLINK〉時間的順序関係ラベルの評価：作業員間の認定傾向の比較

ラベル	DCT	T2E	E2E	MATRIX	全て
関係数	3,839	2,188	2,972	1,245	10,244
after	2352 ∩ 2326 ∩ 2133 = 1961	396 ∩ 441 ∩ 432 = 315	627 ∩ 631 ∩ 639 = 432	292 ∩ 284 ∩ 277 = 198	3667 ∩ 3682 ∩ 3481 = 2906
met-by	0 ∩ 0 ∩ 0 = 0	5 ∩ 10 ∩ 2 = 2	18 ∩ 12 ∩ 3 = 2	7 ∩ 3 ∩ 2 = 1	30 ∩ 25 ∩ 7 = 5
overlapped-by	11 ∩ 5 ∩ 4 = 2	59 ∩ 52 ∩ 42 = 20	3 ∩ 3 ∩ 2 = 0	0 ∩ 0 ∩ 1 = 0	73 ∩ 60 ∩ 49 = 22
finishes	2 ∩ 8 ∩ 1 = 0	10 ∩ 1 ∩ 11 = 0	5 ∩ 8 ∩ 5 = 1	1 ∩ 0 ∩ 0 = 0	18 ∩ 17 ∩ 17 = 1
during	449 ∩ 424 ∩ 650 = 217	105 ∩ 100 ∩ 113 = 62	206 ∩ 139 ∩ 225 = 67	112 ∩ 86 ∩ 134 = 43	872 ∩ 749 ∩ 1122 = 389
started-by	1 ∩ 0 ∩ 0 = 0	9 ∩ 2 ∩ 8 = 0	3 ∩ 14 ∩ 6 = 2	0 ∩ 3 ∩ 0 = 0	13 ∩ 19 ∩ 14 = 2
equal	1 ∩ 17 ∩ 0 = 0	37 ∩ 70 ∩ 51 = 19	263 ∩ 412 ∩ 307 = 154	62 ∩ 140 ∩ 90 = 29	363 ∩ 639 ∩ 448 = 202
starts	2 ∩ 0 ∩ 0 = 0	30 ∩ 9 ∩ 14 = 2	6 ∩ 16 ∩ 2 = 0	0 ∩ 1 ∩ 1 = 0	38 ∩ 26 ∩ 17 = 2
contains	164 ∩ 85 ∩ 144 = 63	830 ∩ 853 ∩ 868 = 671	299 ∩ 292 ∩ 344 = 117	148 ∩ 152 ∩ 188 = 64	1441 ∩ 1382 ∩ 1544 = 915
finished-by	0 ∩ 0 ∩ 0 = 0	3 ∩ 3 ∩ 0 = 0	6 ∩ 7 ∩ 6 = 0	1 ∩ 3 ∩ 0 = 0	10 ∩ 13 ∩ 6 = 0
overlaps	2 ∩ 2 ∩ 4 = 1	75 ∩ 84 ∩ 70 = 32	6 ∩ 27 ∩ 5 = 0	1 ∩ 4 ∩ 3 = 0	84 ∩ 117 ∩ 82 = 33
meets	1 ∩ 13 ∩ 0 = 0	25 ∩ 26 ∩ 2 = 2	88 ∩ 88 ∩ 32 = 22	9 ∩ 15 ∩ 0 = 0	123 ∩ 142 ∩ 34 = 24
before	739 ∩ 767 ∩ 746 = 572	389 ∩ 360 ∩ 383 = 288	1058 ∩ 994 ∩ 1098 = 713	418 ∩ 436 ∩ 422 = 294	2604 ∩ 2557 ∩ 2649 = 1867
is_included	0 ∩ 0 ∩ 0 = 0	0 ∩ 0 ∩ 0 = 0	19 ∩ 2 ∩ 6 = 1	6 ∩ 0 ∩ 1 = 0	25 ∩ 2 ∩ 7 = 1
identity	0 ∩ 0 ∩ 0 = 0	0 ∩ 0 ∩ 1 = 0	11 ∩ 7 ∩ 24 = 2	16 ∩ 5 ∩ 15 = 2	27 ∩ 12 ∩ 40 = 4
includes	0 ∩ 0 ∩ 0 = 0	0 ∩ 0 ∩ 0 = 0	27 ∩ 10 ∩ 2 = 1	18 ∩ 2 ∩ 0 = 0	45 ∩ 12 ∩ 2 = 1
vague	115 ∩ 191 ∩ 157 = 38	212 ∩ 177 ∩ 191 = 100	327 ∩ 309 ∩ 265 = 128	154 ∩ 111 ∩ 111 = 48	808 ∩ 788 ∩ 724 = 314

作業員 A の認定数 ∩ 作業員 B の認定数 ∩ 作業員 C の認定数 = 三者で一致した件数

⁶ “ラベル 13+3+1” および “ラベル 13+1” のラベルと区別するために大文字表記を用いる。

表 9 〈TLINK〉時間的順序関係ラベルの評価：4 種類の関係対ごとの一致率

一致率評価基準	DCT	T2E	E2E	MATRIX	全て
関係数	3,839	2,188	2,972	1,245	10,244
ラベル 13+3+1	0.743 (2,854)	0.691 (1,513)	0.552 (1,642)	0.545 (679)	0.653 (6,688)
ラベル 13+1	0.743 (2,854)	0.691 (1,513)	0.561 (1,667)	0.560 (697)	0.657 (6,731)
ラベル 5+1	0.748 (2,873)	0.734 (1,605)	0.627 (1,862)	0.623 (776)	0.695 (7,116)

三人の作業者の一致率（括弧内は一致したラベル数）

評価基準（ラベル 13+3+1）でも 65.3% の三人の一致率 (Cohen's kappa 0.733) であった。我々の手法では事象構造の認定については複数人で合議的に行い、その後限られた関係について時間的順序関係アノテーションを行っているが、事象構造の認定と関係対に対する関係タグ付与作業を同時に行っている英語のデータ TimeBank 1.2 における〈TLINK〉の一致度（関係対の認定の一致率 55% と一致した関係対に対する関係タグの一致率 77%）と比較しても遜色ないレベルだと考える。4 種類の関係については、“DCT” が最も一致率が高く、次に“T2E” が高かった。これは片方が時間情報表現である場合に、時間情報表現側の時間軸上の絶対位置が推定しやすいことによるからだと考える。

一致率評価基準について始点・終点の境界値一致の認定を緩和することで、“E2E”、“MATRIX” の関係は若干一致率が上がることから、作業者間で事象構造の時間的な境界値にずれが生じていることがわかる。

表 10 に〈EVENT〉の @class ごとの一致率を集計したものを示す。まず、どちらかに静態表現である STATE を含む表現対の作業者間ラベル一致率が低い傾向にある。これは静態表現の始点・終点の認識が作業者間で一致することが困難であることによると考える。左項が時間情報表現 (DCT, TIMEX) であり、右項が STATE である表現をみても、他の時間情報表現-事象表現との関係と比して作業者間ラベル一致率が低い。事象表現を項にとるかどうかの観点でみると、右項が REPORTING, I ACTION の関係が平均よりも高い傾向にある。しかしながら、時間的順序関係が定義されている事象表現対が係り受け構造上の係り受け関係にあるか、また述語項関係になっているかを判断するためには他のアノテーションとの重ね合わせが必要である。今後他機関が作成しているアノテーションを重ね合わせたうえで検討していきたいと考えている。

最後に意味論アノテーションにおける正解のあり方について言及する。テキストが表出する意味レベルの情報の正解は、言語受容者によって完全に復元することは困難であり、100% 正しいものを作成するためには言語生産者によるアノテーション作業が不可欠である。言語生産者によるアノテーション作業を BCCWJ に対して行うことは困難であるため、本研究では作業者三人の結果を統合した形での正解は作成しない。言語受容者の個人の心的空間における時間的順序関係の認識はそれぞれ異なっていてしかるべきであり、受容者ごとに正解があると考えられる。

表 10 <TLINK> 時間的順序関係ラベルの評価: <EVENT>@class ごとの作業者三人の一致率

	DCT	TIMEX	OCC	REP	PER	ASP	LA	LS	STA	ALL
OCCURRENCE	0.739	0.702	0.551	0.625	0.286	0.718	0.559	0.592	0.422	0.656
略記 OCC	(2,352)	(1,320)	(1,602)	(104)	(7)	(39)	(494)	(130)	(102)	(6,159)
REPORTING	0.881	0.697	0.663	0.222	1.000	0.667	0.519	0.368	0.500	0.694
略記 REP	(126)	(66)	(95)	(9)	(2)	(3)	(52)	(19)	(12)	(385)
PERCEPTION	0.815	0.700	0.444	NaN	0.000	NaN	0.500	1.000	0.000	0.646
略記 PER	(27)	(10)	(18)	(0)	(1)	(0)	(6)	(1)	(1)	(65)
ASPECTUAL	0.714	0.615	0.545	1.000	0.000	0.000	0.643	0.000	0.000	0.627
略記 ASP	(63)	(52)	(44)	(6)	(2)	(2)	(14)	(1)	(1)	(185)
LACTION	0.808	0.720	0.576	0.690	0.667	0.765	0.631	0.527	0.333	0.698
略記 LA	(880)	(567)	(491)	(29)	(6)	(17)	(309)	(55)	(51)	(2,407)
LSTATE	0.651	0.686	0.490	0.250	0.750	0.429	0.545	0.875	0.333	0.594
略記 LS	(195)	(86)	(145)	(4)	(4)	(7)	(55)	(16)	(15)	(527)
STATE	0.492	0.398	0.356	0.600	1.000	0.444	0.431	0.333	0.238	0.424
略記 STA	(181)	(83)	(118)	(5)	(3)	(9)	(51)	(9)	(21)	(481)
ALL	0.743	0.691	0.548	0.618	0.560	0.649	0.573	0.562	0.374	0.653
	(3,839)	(2,188)	(2,524)	(157)	(25)	(77)	(984)	(233)	(203)	(10,244)

列が {DCT,TIMEX,<EVENT>@class}, 行が {<EVENT>@class} を表す. 括弧内は各組み合わせの該当件数.

個々の言語受容者の作業結果の正誤判定として, それぞれのアノテーション内での無矛盾性の認定が考えられる. Allen の二次の範囲代数を, 三次以上に拡張すると人の処理能力を超え, 機械的に処理するにも適切な演算が必要になるため, 今後の課題とする⁷.

このアノテーションに基づき解析器の構成を行う場合には何らかの正解を決める必要がある. 正解の設定として, 一人の作業者のアノテーションを正解とする方法, 三人の作業者が一致している部分を正解とする方法, 三人の作業者それぞれの学習モデルを作成し多数決を取る方法などの様々な方法が考えられる. 高性能な構造学習器を構成するためにどのように正解を認めるかについては, 工学研究者に委ねたい.

5 おわりに

本研究では『現代日本語書き言葉均衡コーパス』のコアデータ中の新聞データに対して, 時間的順序関係のアノテーションを行い, アノテーションの一致傾向について報告した. 時間的順序関係を付与する事象表現の認定にあたり, 時間軸上のインスタンスの認定可能性や, 取りうる項が事象である場合に他の事象表現にどのような影響を与えるのかに基づいて, 事象表現

⁷ Allen の範囲代数を拡張すると, 三次で 409 クラス, 四次で 23,917 クラス, 五次で 2,244,361 クラスになることが知られている.

を 7+2 種類に分類した。

次に、三人の作業者による時間的順序関係の一致率などを検討した結果、事象構造の時間軸上の始点・終点の認識は揺れるものの、時間軸上の前後関係は時間情報表現にまつわるもので 73% 以上（ラベル 5+1 評価で DCT 74.8%, T2E 73.4%）、事象表現にまつわるもので 62% 以上（ラベル 5+1 評価で E2E 62.7%, MATRIX 62.3%）の一致率で付与できることがわかった。

本研究におけるアノテーションの評価は、今回策定した基準や作業者で閉じているために限定的である。今後、データを公開⁸し、他機関で同じ部分に付与されているさまざまなアノテーションを重ね合わせ、齟齬や矛盾を分析することでより深い分析が可能になると考えられる。

また、1 節で述べた (b) の意味での目的に応えるために、本データを学習データとして用いた日本語時間的順序関係推定器の開発を今後行っていきたい。英語の時間的順序関係推定器においては、〈MAKEINSTANCE〉タグに付与された テンス・アスペクトの情報が有効な特徴量となる。一方、日本語においては、テンス・アスペクトは、準アスペクト表現を除くと「ル」-「タ」×「テイル」-「テイタ」の二軸の対立しかない。そのうえ「ル」-「タ」の対立は非過去-過去の対立でしかなく、「ル」は定動詞・不定動詞の両方を表現する。このため、形態素解析結果から直接得られるこれらの情報は時間的順序関係推定器の決定的な特徴量とはならない。一方、BCCWJ の当該箇所には、他機関によりモダリティ情報・係り受け構造・述語項構造などが付与されている。これらを重ね合わせることで実用的な時間的順序関係推定器が作成できると考えている。

謝 辞

本研究を行うにあたり、助言いただきました日本 IBM の吉川克正氏、アノテーションに従事していただいた方々に感謝いたします。本研究は文科省科研費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21 世紀の日本語研究の基盤整備」、国語研基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」および国語研「超大規模コーパス構築プロジェクト」によるものです。本論文の一部は The 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27) で発表したものです (Asahara, Yasuda, Konishi, Imada, and Maekawa 2013)。

参考文献

Allen, J. (1983). “Maintaining knowledge about temporal intervals.” *Communications of the ACM*, **26**, pp. 832–843.

⁸ <http://github.com/masayu-a/BCCWJ-Timebank>

- Asahara, M., Yasuda, S., Konishi, H., Imada, M., and Maekawa, K. (2013). “BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text.” In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*.
- Boguraev, B. and Ando, R. K. (2005). “TimeML-Compliant Text Analysis for Temporal Reasoning.” In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 997–1003.
- Boguraev, B. and Ando, R. K. (2006). “Analysis of TimeBank as a Resource for TimeML parsing.” In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-06)*, pp. 71–76.
- DARPA TIDES (2004). *The TERN evaluation plan; time expression recognition and normalization*. Working papers, TERN Evaluation Workshop.
- Grishman, R. and Sundheim, B. (1996). “Message Understanding Conference-6: a brief history.” In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, pp. 466–471.
- 橋本泰一, 中村俊一 (2010). 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.
- IREX 実行委員会 (1999). IREX ワークショップ予稿集.
- 国立国語研究所 (2011). 『現代日本語書き言葉均衡コーパス』利用の手引き (第 1.0 版).
- 小西光, 浅原正幸, 前川喜久雄 (2013). 『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. 自然言語処理, **20** (2), pp. 201–222.
- 工藤真由美 (1995). アスペクト・テンス体系とテキスト—現代日本語の時間の表現—. ひつじ書房.
- 工藤真由美 (2004). 日本語のアスペクト・テンス・ムード体系—標準語研究を超えて—. ひつじ書房.
- Mani, I. (2006). “Machine Learning of Temporal Relations.” In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL-2006)*, pp. 753–760.
- 中村ちどり (2001). 日本語の時間表現. くろしお出版.
- Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., and Katz, G. (2003a). “TimeML: Robust Specification of Event and Temporal Expressions in Text.” In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, pp. 337–353.
- Pustejovsky, J., Hanks, P., Saurí, R., See, A., Gaizauskas, R., Setzer, A., Sundheim, B., Ferro, L., Lazo, M., Mani, I., and Radev, D. (2003b). “The TIMEBANK Corpus.” In *Proceedings of Corpus Linguistics 2003*, pp. 647–656.

- Pustejovsky, J. and Stubbs, A. (2012). *Natural Language Annotation*. O'Reilly.
- Sekine, S., Sudo, K., and Nobata, C. (2002). "Extended Named Entity Hierarchy." In *Proceeding of the third International Conference on Language Resources Evaluation (LREC-02)*, pp. 1818–1824.
- Setzer, A. (2001). *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study*. Ph.D. thesis, University of Sheffield.
- UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., and Pustejovsky, J. (2013). "SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations." In *2nd Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 1–9, Atlanta, Georgia, USA. Association for Computational Linguistics.
- Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Kats, G., and Pustejovsky, J. (2007). "SemEval-2007 Task 15: TempEval Temporal Relation Identification." In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 75–80.
- Verhagen, M., Saurí, R., Caselli, T., and Pustejovsky, J. (2010). "SemEval-2010 Task 13: TempEval-2." In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, pp. 57–62.

略歴

- 保田 祥：2011 年神戸大学人文学研究科博士後期課程修了。2013 年より国立国語研究所コーパス開発センタープロジェクト PD フェロー。現在に至る。博士（文学）。認知意味論の研究に従事。
- 小西 光：2005 年上智大学文学部卒業。2007 年上智大学文学研究科博士前期課程修了。2008 年より国立国語研究所コーパス開発センタープロジェクト奨励研究員。現在に至る。『日本語書き言葉均衡コーパス』『日本語話し言葉コーパス』『日本語大規模コーパス』の整備に携わる。
- 浅原 正幸：2003 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。2004 年より同大学助教。2012 年より国立国語研究所コーパス開発センター特任准教授。現在に至る。博士（工学）。形式意味論の研究に従事。
- 今田 水穂：2010 年筑波大学人文社会科学研究科博士課程修了。筑波大学特任研究員を経て、2013 年より国立国語研究所コーパス開発センタープロジェクト PD フェロー。現在に至る。博士（言語学）。概念意味論の研究に従事。
- 前川喜久雄：1956 年生。1984 年上智大学大学院外国語学研究科博士後期課程（言語学）中途退学。国立国語研究所教授。言語資源系長。コーパス開発セン

ター長, 副所長, 博士 (学術), 専門は音声学ならびに言語資源学.

(2013 年 4 月 24 日 受付)

(2013 年 7 月 5 日 再受付)

(2013 年 8 月 29 日 採録)