事象に対する網羅的な時間情報アノテーションとその分析

坂口 智洋[†]・河原 大輔[†]・黒橋 禎夫[†]

テキスト中には過去・現在・未来における様々な事象が記述されており、その内容を理解するためにはテキスト中の時間情報を正確に解釈する必要がある。これまで、事象情報と時間情報を関連付けたコーパスが構築されてきたが、これらは開始と終了が比較的明確な事象に着目したものであった。本研究では、網羅的かつ表現力豊かな時間情報アノテーション基準を導入し、京都大学テキストコーパス中の113文書に対するアノテーションとその分析を行った。同コーパスには既に述語項関係や共参照関係のアノテーションがなされており、本アノテーションと合わせてテキスト中の事象・エンティティ・時間を対象とした統合的な時間情報解析に活用することが可能となった。

キーワード：時間情報, 時間軸, アノテーション

Comprehensive Annotation and Analysis of Temporal Information of Events

TOMOHIRO SAKAGUCHI[†], DAISUKE KAWAHARA[†] and SADAO KUHASHI[†]

Various past, present, and future events are described in text. To understand such text, correct interpretation of temporal information is essential. For this purpose, many corpora associating events with temporal information have been constructed. Although these corpora focus on expressions with strong temporality, many expressions have weak temporality, but are still clues to understanding temporal information. In this article, we propose an annotation scheme that comprehensively anchors textual expressions to the time axis. Using this scheme, we annotated 113 documents in the Kyoto University Text Corpus. Because the corpus has already been annotated with predicate-argument structures and coreference relations, it can now be utilized for integrated information analysis of events, entities, and time.

Key Words: *Temporal Information Annotation, Temporal Anchoring*

1 はじめに

Web上では日々多くのテキスト情報が発信されており、これまでに膨大な量のテキストが蓄積されている。この大量のテキストから、あるトピックについての知識を抽出するためには、関

[†] 京都大学大学院情報学研究科, Graduate School of Informatics, Kyoto University

連するテキストの統合・要約・比較を行う情報分析技術が必要である。異なる時期に書かれたテキストや異なる時期について言及しているテキストを対象として分析を行うためには、テキストに含意されている時間情報を正しく解釈する必要がある。これまでに事象情報と時間情報の関係性という観点から多くの研究やタスクが行われてきた。例えば TempEval 1, 2, 3 では、事象－事象表現間、事象－時間表現間の時間的順序関係の推定が行われた (Verhagen, Gaizauskas, Schilder, Hepple, Katz, and Pustejovsky 2007; Verhagen, Sauri, Caselli, and Pustejovsky 2010; UzZaman, Llorens, Derczynski, Allen, Verhagen, and Pustejovsky 2013)。また、SemEval 15 では複数のテキストから事象表現を抽出し、時系列に配置するタイムライン生成タスクが扱われた (Minard, Speranza, Agirre, Aldabe, van Erp, Magnini, Rigau, and Urizar 2015)。

このようなタスクにおいてモデルの学習やシステムの評価を行うため、テキスト中の事象情報と時間情報を関連付けたコーパスが作られてきた (Pustejovsky, Hanks, Sauri, See, Gaizauskas, Setzer, Radev, Beth Sundheim, Ferro, and Lazo 2003; Cassidy, McDowell, Chambers, and Bethard 2014; Reimers, Dehghani, and Gurevych 2016)。これらのコーパスでは開始・終了時が比較的明確な事象表現を対象にアノテーションが行われたが、テキストの時間情報理解のための手掛かりはこれにとどまらない。本研究では、時間性が曖昧な表現を含めた、テキスト中の様々な表現がもつ時間情報を表現力豊かにアノテーションするための基準を提案する。

先行研究における時間情報アノテーションのアプローチは2つに大別される。1つは事象間の時間的順序関係を付与する**相対的なアノテーション方法**である。もう1つは各事象を時間軸に対応させる**絶対的なアノテーション方法**である。前者は小説など時間情報の少ないテキストであっても情報量の多いアノテーションが可能である。後者は新聞などの時間情報の多いテキストにおいて少ないアノテーション量で正確に時間情報を表現できる。本研究は後者のアプローチを発展させるものであり、構築するコーパスはタイムライン生成など時間軸を用いてテキストの比較・統合を行うタスクにおいて学習／評価データとして使用することが可能である。

本アノテーション基準の特徴は次の2つである。1つは、時間性を持ち得る幅広い表現をアノテーション対象とすることである。多くの先行研究は、TimeML (Sauri, Littman, Gaizauskas, Setzer, and Pustejovsky 2006) のガイドラインに従い、何か起きたことやその状態を表す一時性の強い表現である“event”に対してアノテーションを行っている。そのため次の例の「出現しており」のような一時性の弱い表現にはアノテーションが行われない。

(1) インターネット上では様々な事業が急速に**出現**しており、政府でさえ把握できていない。

しかし、一時性の弱い表現がもつ時間情報もテキスト解釈の手掛かりとなり得る。この例の場合、「出現しており」が数年前から現在にかけての事象であるという時間情報をアノテーションすることも重要である。そこで本研究では先行研究より対象を広げ、テキスト中で時間性を持ち得る全ての事象表現、すなわちテキスト中の述語または事態性名詞（サ変名詞、動詞連用

形の名詞化, 形容動詞語幹)を含む基本句全て(以降, 対象表現と呼ぶ)をアノテーション対象とする。ここで基本句とは, 京都大学テキストコーパスで定義されている単位で, 自立語とそれに続く付属語のことである。

本アノテーション基準のもう1つの特徴は, 頻度や期間などの多様な時間情報を扱えることである。(Reimers et al. 2016)は事象の起きる期間を開始点と終了点を用いて時間軸に対応付けたが, 次の例で太字で示す「飛び飛びな時間」や「大きな区間の中のある一部の期間」に起きる事象を正確に時間軸に対応付けることはできなかった。

- (2) 毎週日曜は球場で野球を見る。
- (3) 来週は3日間京都に出張する。
- (4) 昔はよく一緒に遊んでいた。

本研究では, テキスト中に含まれる多様な時間情報をより正確に時間軸に対応付けられる時間タグを導入する。

多様な表現に対して表現力豊かにタグ付けすることで, 個人のテキスト解釈や常識がタグの揺れとして現れる。本研究ではこのような揺れも時間がどのように解釈されているかを知る上で重要だと考えているため, 最終的に複数のアノテータの付与した時間タグを1つに統合することはしない。代わりに, 解釈の違いを尊重しつつ明らかなアノテーションミスのみを修正するアノテーション方法を導入する。

本アノテーション基準を用いて, 京都大学テキストコーパス中の113文書4,534対象表現に対してアノテーションを行った。その結果, 対象表現の76%に時間性が認められ, そのうち35%(全体の26%)で本稿で新たに提案する記法が用いられた。同コーパスには, 既に述語項関係や共参照関係のアノテーションがなされているため, 本アノテーションと合わせてテキスト中の事象・エンティティ・時間を対象とした統合的な時間情報解析に活用することが可能となる。

2 関連研究

事象情報と時間情報を関連付けたコーパスは, これまでにも多く作られており, これらのアノテーション方法は大きく2つのアプローチに分けられる。1つは, 事象間の時間的順序関係を付与する方法である。事象情報や時間情報を扱う多くのタスクで利用されているTimeBank Corpus (Pustejovsky et al. 2003)には, TimeMLの基準に基づいて事象・時間表現情報がアノテーションされており, さらに時間的順序関係を表すTLINK, 事象間の関係を表すSLINK, 相動詞と事象の関係を表すALINKの3つの関係が付与されている。当初はアノテータが重要と判断した表現間のみアノテーションがなされていたためスパースであったが, 後のTempEvalタスクはこれを発展させ, 同一文中と隣接文間の関係に対してアノテーションを行った。BCCWJ-

TimeBank (Asahara, Kato, Konishi, Imada, and Maekawa 2014) は、現代日本語書き言葉均衡コーパス (BCCWJ) の新聞記事に対して TimeBank Corpus に準拠した基準でアノテーションを行ったものである。

このような時間的順序関係をより密にアノテーションしたコーパスも存在する。(Kolomiyets, Bethard, and Moens 2012) は子ども向けの物語コーパス中の各事象表現に対してその最も近い事象表現との時間的順序関係をアノテーションした。また TimeBank Dense Corpus (Cassidy et al. 2014) は、同一文内・隣接文間の全ての事象－時間表現間、事象－事象表現間に対してアノテーションを行っている。

もう1つのアプローチは、事象を時間軸に対応させる方法である。EventStatus Corpus (Huang, Cases, Jurafsky, Condoravdi, and Riloff 2016) は、社会不安に関する新聞記事中の事象表現に対して *Past*, *On-going*, *Future Planned*, *Future Alert*, *Future Possible* の5つのラベルをアノテーションした。(Asakura, Hangyo, and Komachi 2016) は、ソーシャルメディアに投稿された洪水に関するテキストに含まれる事象表現に対して *PAST*, *PRESENT*, *FUTURE* のラベルと事実性に関する3つの値 (*high probability*, *low probability*, *unmentioned*) をアノテーションした。

より細かい粒度で時間軸に対応付けたコーパスも存在する。(Reimers et al. 2016) は、TimeBank Corpus 中の事象表現に対して時間値を付与した。時間値は日にちの粒度で付与されており、事象が1日内で終わるものか、複数日に跨るものかでタグが異なる。前者にはその事象が起きた日付を、後者にはその事象の開始日と終了日を付与する。次の例文の「出発した」は1日内の事象であり、この方法でアノテーションする場合は 1980-05-26 を付与する。「過ごした」は複数日に跨る事象であるため、*beginPoint=1980-05-26 endPoint=1980-06-01* のように開始日と終了日を付与する。

(5) 1980年5月26日、彼は宇宙に向けて**出発**した。サリュート6号で6日間を**過ご**した。

事象の正確な日付が分からない場合は、*after* と *before* を用いて記述する。次の例文の「滞在した」には *beginPoint=after 1984-10-01 before 1984-10-31 endPoint=after 1984-10-01 before 1984-10-31*, 「選ばれた」には *after 2014-01-01 before 2014-12-31* が付与される。

(6) サリバンは1984年10月、チャレンジャー号のメンバーとして宇宙に**滞在**した。2014年にはタイム100に**選ば**れた。

また、時間性をもたない事象に対しては、*n/a* を付与した。

彼らのアノテーションでは、全事象の約6割が1日内の事象、約4割が複数日に跨る事象である。前者のうち日付が明確なものは56%、後者のうち開始日が明確なものは20%、終了日が明確なものは16%であり、全事象の64%が *after* または *before* を用いて表されている。時間性をもたない事象は全体の0.7%であった。

本稿ではこの事象を時間軸に対応させるアプローチを発展させ、テキスト中の多様な時間情報に対応できるタグ付け基準を提案する。

3 アノテーション基準

1章で述べたように、本研究ではテキスト中の述語または事態性名詞（サ変名詞、動詞連用形の名詞化、形容動詞語幹）を含む基本句全てを対象表現として、それらに時間タグを付与する。ここで基本句とは、自立語とそれに続く付属語のことである。基本句を基本単位とすることにより、「行くつもりだ」や「勝てたかもしれない」のような助詞や接尾辞を含む動詞句をひとかたまりとして時間性を考えることができる。英語を対象とした先行研究の多くが、開始と終了が比較的明瞭で一時的の強い事象である、TimeMLにおける“event”をアノテーション対象とするのに対し、本研究では時間性をもち得るより幅広いこれらの表現をアノテーション対象とする。対象表現の例を次に示す。例(7)の「行くつもりだ」は動詞を含む基本句であるため、また例(8)の「所属」は事態性名詞であるため対象表現である。例(9)には、事態性名詞の基本句である「結婚を」と、動詞を含む基本句である「考えたい」の2つの対象表現が存在する。

- (7) 明日京都に行くつもりだ。
- (8) 連合所属議員
- (9) そろそろ 結婚を 考えたい。

アノテータには、本章で述べるアノテーション基準に従って、対象表現に対するアノテーションを依頼した。アノテータは、まず対象表現が時間性をもつかどうかを判定する。時間性をもつ場合はテキストの作成日 (Document Creation Time, DCT) と文脈を考慮して対応する時間タグを付与する。時間性を持たない場合は *t:n/a (not applicable)* という時間タグを記す。

時間性をもつ場合、時間タグは時間基本単位 (Time Base Unit, TBU) またはその組み合わせとして表す (表1)。時間基本単位とは、特定の時点や期間を表すもので、5種類のタグを定義する。さらに、時間基本単位中の一部の期間や繰り返しを表す3通りの方法を導入し、多様な時間情報を表現する。(Reimers et al. 2016) が扱った時間情報は、表1中の1, 3, b, *t:n/a* である。

時間タグは、先行研究と同様、日にちを最小粒度とする。これは、情報分析において着目したい粒度が数日から数年であることが多いことによる。例えば次の文のように、2017年4月29日に書かれたテキストでは、「帰った」はその日の18時のことであるが、日にち以下の粒度の情報は捨てて2017年4月29日という情報をタグ付けする。

- (10) [DCT: 2017-04-29] 今日は夜6時に帰った。

表 1 時間タグの一覧

時間性あり	時間基本単位 (Time Base Unit, TBU)	1. 日付タグ (日・月・年) (例) <i>t:1995-01-05</i>
		2*. 漠然とした時間タグ (過去・現在・未来) (例) <i>t:PRESENT</i>
		3. 期間を表す時間タグ (開始日 ~ 終了日) (例) <i>t:1995-01-05~1995-01-07</i>
		4*. 相対的な時間タグ (時間共参照) (例) <i>t:選挙の</i>
		5*. 発話日を表す時間タグ (例) <i>t:UD+P1D</i>
時間性なし	TBU の一部分	a*. TBU 中の特定の幅をもつ期間 (span) (例) <i>t:1995-01,span:P1W</i>
		b. TBU 中の不特定の期間 (span:part) (例) <i>t:1995-01,span:part</i>
		c*. TBU 中の繰り返し (freq) (例) <i>t:1995-01,freq:2/P1W</i>
		(例) <i>t:n/a</i>

*は本研究で新しく導入したタグを表す.

3.1 対象表現の時間性判定

対象表現が時間性をもつかどうかは, 過去から未来の間において表現の表す動作や状態に変化があるか否かで判断する. 文脈によって変化の有無や度合いの解釈は変わり得るため, 次のような事例を共通認識としつつ, 具体的な判断基準は各アノテータに委ねた.

時間性をもつ例をいくつか挙げる.

- (11) 明日京都に行く。
- (12) 言語処理研究が盛んだ。
- (13) あの子は背が低かった。

例 (11) の「行く」は, 明日という特定の日に起きるものであるため, 時間性をもつ. 例 (12) の「盛んだ」は, いつからいつまでかは分からないがある限られた時期のことであるため時間性をもつと考える. 例 (13) の「低かった」も, 現在はそうではないことを示唆しているため, 時間性をもつと考える. これらの例において, 先行研究がアノテーション対象としているのは, 例 (11) のみである.

次に時間性をもたない例を示す.

- (14) ウサギは草を食べる 動物だ。

(15) 彼の目は黒い。

例 (14) の「食べる」と「動物だ」は、昔から変わらない一般的な事柄であるため時間性をもたないとする。例 (15) の「黒い」も同様である。ただし、次の文のように現在は異なることを示唆する表現の場合は時間性をもつと解釈する。

(16) 以前は目の色が黒かった。

3.2 時間基本単位

3.2.1 日付・期間を表すタグ

日付の時間情報は、 t タグにその時間の表す値を記すことで表現する。 $t:YYYY$ や $t:YYYY-MM-DD$ など、BCCWJ-TimeBank で定義されている時間表現の時間値の記法で記す。次の例の「到着した」は 2017 年 4 月 28 日の出来事であるため、時間値 $t:2017-04-28$ を付与する。

(17) [DCT: 2017-04-29] 昨日大統領がニューヨークに到着した。

→ $t:2017-04-28$

本研究では先行研究とは異なり、日にちより大きな粒度の時間タグを許す。例えば次の例の「暑かった」には、 $t:2016-08$ を付与する。このタグは、必ずしも厳密に 2016 年 8 月 1 日から 31 日までの期間を表すわけではない。「8 月」という表現は「8 月 1 日から 31 日まで」という表現と比べ、その表す期間は漠然と捉えられる。本研究におけるタグの粒度はこのような漠然性を含意する。

(18) [DCT: 2017-04-29] 今年の 8 月は暑かった。

→ $t:2016-08$

アノテータの負担軽減のため、次のような省略表記を導入する。

- 文書作成日時の日付は $t:DCT$ と記すことができる。
- ある日から一定期間前／後の日を引き算／足し算で記すことができる。この際、期間は BCCWJ-TimeBank で定義された期間表現の時間値の記法を用いる。例えば、1 年間は $P1Y$ 、1 ヶ月は $P1M$ 、1 週間は $P1W$ 、1 日間は $P1D$ と表す。次の例の「行く」の時間値は $t:DCT+P1W$ 、「行った」の時間値は $t:DCT-P1W$ と表せる。

(19) 私も来週そこに行く。

→ $t:DCT+P1W$

(20) 私も先週そこに行った。

→ $t:DCT-P1W$

3.2.2 漠然とした時間を表すタグ

テキスト中には漠然とした時間情報を表す表現も多く存在する。例えば次の文では、「住んでいた」が過去のいつ、どのくらいの期間のことであるのか具体的には分からない。

(21) 昔広島に住んでいた。

(Reimers et al. 2016) はこれを「今日までのある日から、今日までのある日まで」と捉え、*beginPoint=before DCT endPoint=before DCT* とタグ付けした。本研究では、このような漠然とした時間情報をより正確に表現するため、新たなタグを導入する。

文書作成日を基準として、漠然とした過去、現在、未来は、それぞれ *t:PAST*, *t:PRESENT*, *t:FUTURE* と記す。ここで、「現在」は文書作成日の少し前から少し後までを表す。例えば次の文の「持ち込める」は、文書作成日だけでなく、文書作成日から少し前や後でも成り立つことだと考えられるので、*t:PRESENT* を付与する。

(22) 国内線では飲み物を持ち込める。

→ *t:PRESENT*

過去と未来については、それぞれ時間的距離に応じて *t:PAST-M* と *t:PAST-Y*, *t:FUTURE-M* と *t:FUTURE-Y* という表記を導入する。*t:PAST-M* は数ヶ月前を、*t:PAST-Y* は数年前を表す¹。数年以上前、あるいはどのくらい昔なのかが不明な場合は *t:PAST* を用いる。未来についても同様である。

漠然とした時間表現はこれ以外にもある。「1980 年頃」や「約 3 年間」など数値の曖昧性が明示されている表現の場合、その曖昧な数値の直後に *ap* (*approximately*) と記す。次の例の「建てられた」には *t:1980ap* を付与する。

(23) 1980 年頃に建てられた建物

→ *t:1980ap*

3.2.3 開始・終了時を用いて期間を表す方法

期間の時間値は、開始時と終了時を ~ で結ぶことで表す。これは (Reimers et al. 2016) の *beginPoint*, *endPoint* に対応する。例 (5) の「過ごした」には *t:1980-05-26~1980-06-01* を付与する。

(5) 1980 年 5 月 26 日、彼は宇宙に向けて出発した。サリュート 6 号で 6 日間を過ごした。

→ *t:1980-05-26~1980-06-01*

¹ 数週間前までの過去は、3.2.3 節の ~ を用いて *t:~DCT* と表す。

事象の開始時, 終了時のいずれかが不明かつ近い過去／未来である場合はこれを書かない. 次の例文の「忙しかった」の時間タグは $t:\sim 2017-04-28$ となる.

(24) [DCT: 2017-04-29] 昨日まで忙しかった。

→ $t:\sim 2017-04-28$

事象の開始時, 終了時が遠い過去や未来である場合は, *PAST* や *FUTURE* を用いる.

3.2.4 相対的な時間を表すタグ

事象を時間軸に対応付けるアプローチの弱点の1つは, 小説など時間表現が少ないテキストでは多くの事象が時間軸と対応付けられない可能性があることである. このような場合, 事象表現間の相対的な時間関係をタグ付けするアプローチの方が情報量の多いアノテーションが可能である. そこで本研究では, 絶対的な時間値が分からない場合は相対的な時間値を付与する.

対象表現の具体的な日付が分からないが, 同一文中の他の基本句との時間関係が分かる場合, これを時間値として記す (時間共参照と呼ぶ). 例えば次の文の「起きた」は, 具体的な日付は分からないが基本句「選挙の」の表す日の翌日であることは分かる. この場合, $t:\text{選挙の}+P1D$ と記す.

(25) 選挙の翌日、大規模なデモが起きた。

→ $t:\text{選挙の}+P1D$

参照できる基本句が複数ある場合は, 1. 絶対的な時間値が付与されている基本句, 2. 距離が近い基本句の順で優先順位を付け, 最も優先順位の高いもの1つを選択する.

3.2.5 発話日時を表すタグ

会話文やインタビューなどでは, その発話日時が不明なことが多い. 文脈から具体的な日付が分かる場合はそれを利用してタグ付けを行うが, 分からない場合は発話日を $t:UD$ (*Utterance Day*) として記述する. 例えば次の文の「頑張るしかない」には $t:UD+P1D$ を付与する.

(26) 「明日頑張るしかない」と監督は言った。

→ $t:UD+P1D$

ただし, 上の文の「言った」のように, 発話の外の表現には *UD* は用いず, 絶対的な時間値を記す.

3.3 時間基本単位 (TBU) の一部分の表現

3.3.1 TBU 内の一部の期間を表す方法

ある大きな TBU の中の一部の期間は、大きい期間を表す *t* タグと小さい期間を表す *span* タグを組み合わせることで表現する。 *span* タグは、期間の長さが分かる場合は BCCWJ-TimeBank で定義された期間表現の時間値の記法を用いて記す。例えば 3 日間は *span:P3D*, 3 週間は *span:P3W*, 3 年間は *span:P3Y* と表す。期間の長さが分からない場合は *span:part* と記す。例 (6) の場合、「滞在した」は 1984 年 10 月のある期間なので *t:1984-10,span:part* を、「選ばれた」は 2014 年のある日のことなので *t:2014,span:P1D* を付与する。

- (6) サリバンは 1984 年 10 月、チャレンジャー号のメンバーとして宇宙に滞在した。2014 年にはタイム 100 に選ばれた。

→ 「滞在した」 *t:1984-10,span:part*
「選ばれた」 *t:2014,span:P1D*

(Reimers et al. 2016) と比較すると、*span:part* は彼らの *before*, *after* を用いた記法に対応する。

3.3.2 TBU 中の繰り返しを表す方法

対象表現は常に連続した期間として表せるわけではない。「毎週日曜日」に行う事象や「3 日に 1 回」行う事象もある。このように飛び飛びで複数日に渡って起きる事象は、*t* タグや *span* タグに加えて繰り返し規則を表す *freq* タグを用いて表す。

繰り返しに関するアノテーション方法は次の 3 つがある。

- 「週に 2 回」や「3 日に 1 度」のように、繰り返し規則が一定の期間中に起きた回数として表される場合、回数／期間を *freq* タグに記す。次の例文の「通っている」には *t:2016-07~DCT,freq:2/P1W* とタグ付けする。

- (27) [DCT: 2017-04-29] 昨年 7 月から週に 2 回プールに通っている。

→ *t:2016-07~DCT,freq:2/P1W*

- 「毎月 25 日」や「毎週日曜日」のように、繰り返し規則が特定の日付や曜日として表される場合、その日付や曜日を *freq* タグに記す。このとき、BCCWJ-TimeBank のタグ付け基準を拡張し、いかなる数字も入るという意味で YYYY-MM-DD の各部分に@を入れることを許す。例えば *freq:@@@@-@@-25* は毎月 25 日を表す。次に例を挙げる。

- (28) 骨董市は毎月 25 日に開催される。

→ *t:PRESENT,freq:@@@@-@@-25*

- (29) 毎週日曜はプールに行く。

→ *t:PRESENT,freq:@@@@-@@-Sun*

- 具体的な回数や頻度が文脈から分らない場合, 頻度を抽象的に表した 4 つの値, *usually, often, sometimes, rarely* のいずれかを用いる. 次の例文の「行く」には *t:PRESENT, freq:sometimes* を付与する.

(30) [DCT: 2017-04-29] スターバックスに時々行く。

→ *t:PRESENT,freq:sometimes*

4 アノテーション結果

4.1 アノテーション対象

上述のアノテーション基準を用いて, 京都大学テキストコーパス (Kawahara, Kurohashi, and Hasida 2002) 中の一部の記事に対してアノテーションを行った. 同コーパスは, 1995 年 1 月 1 日から 17 日までの毎日新聞の記事に各種言語情報を人手で付与したものである. この中から当時話題になっていたトピックを 11 個選定し, トピックに関連する 113 記事 856 文, 4,534 表現にアノテーションを行った (表 2). 4,534 表現のうち, 述語は 3,072 表現, 事態性名詞は 1,462 表現であった.

4.2 アノテーション方法

3 名のアノテータにアノテーションを依頼した. 本研究では個人の感覚や常識によって解釈が変わるような表現も対象としているため, アノテータの付与したタグを最終的に 1 つに統合

表 2 アノテーションを行った記事数の分布

トピック	文書作成日 (いずれも 1995 年)						
	1 月 1 日	1 月 3 日	1 月 4 日	1 月 5 日	1 月 6 日	1 月 7 日	計
チェチェン	1	3	5	7	7	4	27
新党結成	2	0	0	4	4	7	17
アメリカ動向	2	1	2	2	3	2	12
首相動向	1	2	4	3	1	0	11
アメリカズカップ	10	0	0	0	0	0	10
中国動向	4	0	1	4	0	0	9
メコン川開発	7	0	0	0	0	0	7
長野での雪崩	0	0	0	4	2	0	6
関西空港	1	0	0	4	1	0	6
ロシア動向	1	1	1	0	1	1	5
メキシコ経済	0	0	2	1	0	0	3

図 1 3 名のアノテータによるアノテーション方法. データを 3 分割し, 各アノテータは第 1 段階でそのうちの 2 つ, 第 2 段階で残りの 1 つを担当する. 第 2 段階では他の 2 人が第 1 段階で付与したタグを見ることができる.

することはしない. 代わりに, 他のアノテータの解釈を尊重しつつ, 明らかなアノテーションミスは修正する 2 段階のアノテーションを行う. まず対象となる文書群全体を 3 つに分割し, 各アノテータは第 1 段階でそのうちの 2 つ, 第 2 段階で残りの 1 つを担当する (図 1). これにより, 最終的に全員が各文書に 1 度ずつアノテーションを行う. 第 1 段階では各アノテータが独立にアノテーションを行うのに対し, 第 2 段階では他の 2 人が第 1 段階で付与したタグを見た上で同じタグもしくは独自のタグを付与する. 第 2 段階では, もし既に付けられたタグに明らかな誤りがある場合は印をつける. 印を付けられたタグは全体の 2% であり, 本稿における分析では欠損値として扱う.

4.3 時間タグの分布

アノテーションされた全時間タグの分布を表 3 に示す. 約 25% の対象表現が時間性をもたない判定された. 日付タグが全体の約 25%, 期間を表すタグが約 15% を占める一方で, 漠然とした時間タグは約 10%, 時間共参照を含むタグは数% と少ない. 新聞を対象にタグ付けを行ったため, 表現の多くが時間軸に対応付けられたと考えられる. また, 繰り返しを表す *freq* タグは全体の 1% とほとんど表れなかった. 本稿で新たに提案したタグは, 全体の約 25% を占めた.

表 4 は, アノテーションの第 2 段階における, 述語と事態性名詞それぞれに付与された時間タグの分布を示したものである. 述語に付与されたタグの 27% を日付タグが占めるのに対し, 事態性名詞では約半分の 12% である. 一方, 事態性名詞では時間性なしと判定された表現が多く, 述語の約 2 倍の割合に当たる 35% を占める. これは, 「全欧安保協力機構」「地方旅行の自由化」など, 事態性名詞が組織や一般的な事象などを表すことも多いためである.

4.4 アノテータ間一致率

Krippendorff's α (Krippendorff 2004; Hayes and Krippendorff 2007) を用いてアノテータ間一致率を算出した. (Reimers et al. 2016) と同様の 2 つの基準を用いた. 1 つはタグの一致度を厳格に測る基準 (Strict 基準) で, 時間タグが完全に一致するか否かを判定する. 例えば *t:1994-12-31* は *t:1994-12-31* と一致するが, *t:~1994-12-31* とは一致しない. もう 1 つは時間タグの部分一致

表 3 第 1, 第 2 段階でアノテーションされた全ての時間タグの分布

	アノテータ 1	アノテータ 2	アノテータ 3	平均
1. 日付タグ	1,195 (26.4%)	1,145 (25.3%)	938 (20.7%)	1,093 (24.1%)
年	35 (0.8%)	47 (1.0%)	16 (0.4%)	33 (0.7%)
月	9 (0.2%)	14 (0.3%)	3 (0.1%)	9 (0.2%)
日	1,151 (25.4%)	1,084 (23.9%)	919 (20.3%)	1,051 (23.2%)
2*. 漠然とした時間タグ	617 (13.6%)	375 (8.3%)	249 (5.5%)	414 (9.1%)
<i>t:PRESENT</i>	520 (11.5%)	257 (5.7%)	195 (4.3%)	324 (7.2%)
<i>t:PAST</i>	40 (0.9%)	25 (0.6%)	17 (0.4%)	27 (0.6%)
<i>t:FUTURE</i>	57 (1.3%)	89 (2.0%)	31 (0.7%)	59 (1.3%)
<i>t:ap</i>	0 (0.0%)	4 (0.1%)	6 (0.1%)	3 (0.1%)
3. 期間を表す時間タグ (～)	387 (8.5%)	562 (12.4%)	842 (18.6%)	597 (13.2%)
4*. 相対的な時間 (時間共参照) を含むタグ	138 (3.0%)	58 (1.3%)	207 (4.6%)	134 (3.0%)
5*. 発話日 (<i>t:UD</i>) を含むタグ	77 (1.7%)	77 (1.7%)	106 (2.3%)	87 (1.9%)
a*. TBU 中の特定の幅をもつ期間 (span)	540 (11.9%)	447 (9.9%)	550 (12.1%)	512 (11.3%)
日付タグ + span	46 (1.0%)	69 (1.5%)	96 (2.1%)	70 (1.6%)
～ + span	482 (10.6%)	357 (7.9%)	434 (9.6%)	424 (9.4%)
漠然タグ + span	12 (0.3%)	21 (0.5%)	20 (0.4%)	18 (0.4%)
b. TBU 中の不特定の期間 (span:part)	455 (10.0%)	561 (12.4%)	478 (10.5%)	498 (11.0%)
日付タグ + span:part	36 (0.8%)	56 (1.2%)	46 (1.0%)	46 (1.0%)
～ + span:part	373 (8.2%)	475 (10.5%)	391 (8.6%)	413 (9.1%)
漠然タグ + span:part	46 (1.0%)	30 (0.7%)	41 (0.9%)	39 (0.9%)
c*. TBU 中の繰り返し (freq)	46 (1.0%)	52 (1.2%)	47 (1.0%)	48 (1.1%)
時間性をもたないタグ (<i>t:n/a</i>)	1,071 (23.6%)	1,077 (23.8%)	1,060 (23.4%)	1,069 (23.6%)
第 2 段階でチェックを付けられたタグ	8 (0.2%)	180 (4.0%)	57 (1.3%)	82 (1.8%)
合計	4,534	4,534	4,534	4,534
本研究で新たに提案されたタグ (*の項目)	1,418 (31.3%)	1,009 (22.3%)	1,159 (25.6%)	1,195 (26.4%)

インデントされた項目は内訳を表す。また, *は本研究で新たに導入したタグを表す。

表 4 第 2 段階でアノテーションされた, 述語と事態性名詞に対する時間タグの分布

	述語	事態性名詞	合計
1. 日付タグ	844 (27%)	172 (12%)	1,016 (22%)
2*. 漠然とした時間タグ	272 (9%)	105 (7%)	377 (8%)
3. 期間を表す時間タグ	460 (15%)	210 (14%)	670 (15%)
4*. 相対的な時間タグ (時間共参照)	120 (4%)	42 (3%)	162 (4%)
5*. 発話日を含むタグ	68 (2%)	8 (1%)	76 (2%)
a*. TBU 中の特定の幅をもつ期間 (span)	380 (12%)	173 (12%)	553 (12%)
b. TBU 中の不特定の期間 (span:part)	311 (10%)	216 (15%)	527 (12%)
c*. TBU 中の繰り返し (freq)	39 (1%)	18 (1%)	57 (1%)
時間性なし	578 (19%)	518 (35%)	1,096 (24%)
合計	3,072	1,462	4,534

を認める緩やかな基準 (Relax 基準) である。アノテータ間の時間タグが1日でも重なっていたら一致とし、全く重なっていなかったら不一致とする。例えば $t:1994-12-31$ と $t:\sim 1994-12-31$ は、両者の範囲が重なっているため一致と判定する。一方、 $t:\sim 1994-12-31$ と $t:1995-01-01$ のように、1日も重なっていない場合は不一致と判定する。このとき、 $t:n/a$ と時間共参照については部分一致を認めず、完全に同じでない場合は不一致として扱う。各段階終了時の一致率を表5に記す。表の括弧内の2つの数値はそれぞれ述語、事態性名詞のみを対象とした場合の一致率である。最終時とは、第2段階までのアノテーションを終えた、最終的なタグを表す。「 $t:n/a$ を除く」は、各段階で1人でも $t:n/a$ を付けた表現 (第1段階終了時、最終時ともに約1,300個) を除いたものである。第1段階終了時では2名の、最終時では3名のアノテータ間一致率を算出した。

第1段階終了時と最終時の一致率を比較すると、後者が大幅に上がっている。これは、前者は独立に付与されたタグであるのに対し、後者は他のアノテータのタグを見て付与された第2段階のタグが含まれていることが影響していると考えられる。また、 $t:n/a$ が付与された表現を除いて評価すると、Strict 基準において大幅な一致率向上が見られた。このことから、Relax 基準での一致率が低い原因の1つが、時間性判定の難しさにあることが分かる。

述語と事態性名詞の一致率を比較すると、両基準ともに事態性名詞の一致率が低い。事態性名詞は述語に比べて時間性のない表現の占める割合が多いため、時間性判定の難しさの影響をより大きく受けていると考えられる。 $t:n/a$ が付与された表現を除いても Strict 基準では事態性名詞の一致率が低く、時間が明確な表現は少ないことが分かる。ただし、これらの値は Relax 基準では大幅に上がり、また述語との差もほぼなくなるため、タグは完全には一致しないもののアノテータ間の認識に大きな隔たりはないと考えられる。

(Reimers et al. 2016) の一致率と比較すると、特に Strict 基準において低いものとなっている。これは時間タグのバリエーションを増やしたことにより個人の解釈の揺れが多く反映されたことを表している。

両基準におけるアノテータ間のタグの一致／不一致について、次章でより詳細に述べる。

表 5 Krippendorff's α を用いて算出したアノテータ間一致率

	Strict 基準	Relax 基準
第1段階終了時	0.417 (0.439/0.353)	0.719 (0.743/0.659)
最終時	0.554 (0.571/0.506)	0.802 (0.820/0.756)
第1段階終了時 ($t:n/a$ を除く)	0.380 (0.410/0.293)	0.803 (0.803/0.798)
最終時 ($t:n/a$ を除く)	0.526 (0.548/0.461)	0.867 (0.867/0.865)
[Reimers+ 16]	0.617	0.912

括弧内の数値は (述語／事態性名詞) における一致率を表す。

5 アノテータ間でのタグの揺れ分析

本研究で提案した時間タグは、先行研究と比べ時間情報をより正確に表現できるようになった一方でアノテータの解釈に敏感である。本章では、アノテータ間でどのように時間タグが揺れたかを分析する。

具体的な時間値にとらわれずに時間タグの特徴を扱うため、時間タグを粒度の面から抽象化する。例えば、*t:1994-12-31* や *t:DCT* などの日付を表す時間タグは *DAY*、*t:~1994-12-31* は *~DAY*、*t:1994* は *YEAR* とする。また、*span* タグや *freq* タグはその値を省略して表記する。例えば *t:~1994-12-31,span:PID* や *t:~1994-12-31,span:part* は *~DAY,span* とする。

本章では、アノテータが独立にアノテーションを行った、第1段階のアノテーション結果を対象として分析する。表6、表7は、それぞれ Strict 基準と Relax 基準において、アノテータ間でどのように時間タグが揺れたかをまとめたものである。ここで、アノテーションの一致の判定は元の時間値を用いて行い、集計のみ抽象化を行った値を使用した。

表6を見ると、Strict 基準ではアノテータ間で一致したタグの約7割が *DAY* と *n/a* であり、不一致の多くは *n/a* か否かの判定、あるいは *DAY* と *~DAY* などの日付と期間の解釈の違いに起因するものであることが分かる。表7を見ると、Relax 基準における不一致のほとんどは *n/a* か否かの判定である。Strict 基準で見られた日付と期間の解釈の違いのほとんどはこの基準では一致しており、領域が重なっていたことが分かる。

このような、アノテータ間の時間性判定と日付・期間の解釈の揺れについて次節以降、具体例を通して分析する。

表6 第1段階のアノテーション結果に対する、『Strict 基準』におけるアノテータ間の時間タグの一致／不一致頻度

アノテータ間で一致した時間タグ			アノテータ間で一致しなかった時間タグ		
時間タグペア		頻度	時間タグペア		頻度
<i>n/a</i>	<i>n/a</i>	800	<i>PRESENT</i>	<i>n/a</i>	110
<i>DAY</i>	<i>DAY</i>	740	<i>DAY</i>	<i>n/a</i>	105
<i>~DAY,span</i>	<i>~DAY,span</i>	142	<i>DAY</i>	<i>~DAY,span</i>	104
<i>PRESENT</i>	<i>PRESENT</i>	113	<i>~DAY,span</i>	<i>~DAY,span</i>	77
<i>DAY~,span</i>	<i>DAY~,span</i>	54	<i>DAY</i>	<i>~DAY</i>	73
<i>DAY~DAY</i>	<i>DAY~DAY</i>	38	<i>PRESENT</i>	<i>~DAY,span</i>	59
<i>YEAR</i>	<i>YEAR</i>	12	<i>PRESENT</i>	<i>~DAY</i>	49
<i>DAY~FUTURE</i>	<i>DAY~FUTURE</i>	10	<i>PRESENT</i>	<i>PAST~DAY</i>	49
合計		2,045	合計		2,275

表 7 第 1 段階のアノテーション結果に対する、『Relax 基準』におけるアノテータ間の時間タグの一致／不一致頻度

アノテータ間で一致した時間タグ			アノテータ間で一致しなかった時間タグ		
時間タグペア		頻度	時間タグペア		頻度
n/a	n/a	800	PRESENT	n/a	110
DAY	DAY	741	DAY	n/a	105
~DAY,span	~DAY,span	219	DAY	DAY	44
PRESENT	PRESENT	113	DAY~,span	n/a	36
~DAY,span	DAY	90	~DAY,span	n/a	31
DAY~,span	DAY~,span	85	DAY~FUTURE	n/a	20
DAY	~DAY	67	DAY~FUTURE,span	n/a	19
~DAY,span	PRESENT	59	DAY~DAY	n/a	14
合計		3,533	合計		787

5.1 時間性の判定

Relax 基準において、アノテータ間でタグが一致しない最大の原因は、アノテータによって時間性の判断が揺れることにある。n/a タグと共起しやすいタグは、頻度順に n/a (76.6%), DAY (5.3%), PRESENT (5.0%), ~DAY,span (1.7%) となっており、8 割近くの割合でアノテータ間で n/a が一致し、そうでない場合の約 4 割は片方のアノテータが DAY か PRESENT を付与している。このような表現には状態や役職、組織を表すものが多く、ある程度普遍的なものを見るか、期間としては長くても一時的なものとするかで判断が分かれている。

次の例では、アノテータのタグが *t:PRESENT* と *t:n/a* で揺れた。

(31) 大統領官邸のある中心部

→ *t:PRESENT* vs *t:n/a*

時間性を認めたアノテータは、記事の執筆時前後では大統領官邸が中心部にあるものの、過去、あるいは未来にはそれが変わる可能性があるとして解釈したのに対し、時間性を認めなかったアノテータは、変わる可能性はほとんどない、半永続的なことと解釈したと考えられる。

5.2 日付と期間の解釈

(Reimers et al. 2016) も指摘しているとおり、ある事象がある日一日内のことなのか、複数日に跨ることなのかをテキストから判断するのは難しい。また、ある事象の開始時や終了時を明確化することも容易ではない。このような曖昧性は、本アノテーションにおいては DAY, ~DAY, ~DAY,span, DAY~, DAY~,span, PRESENT の揺れとして現れる。中でも多いのは DAY と ~DAY,span の間の揺れである。特に DAY が DCT の場合が多く、文書作成日に起きたのか、それまでに起きたのかの解釈が難しいことを表している。

これはアノテーション対象が新聞記事であることが要因の1つであると考えられる。次の例では、アノテータのタグが $t:DCT$ と $t:\sim DCT, span:part$ で揺れた。前者はこの表現を文書作成日のことと解釈したのに対し、後者はより長い期間であると解釈したと考えられる。

(32) しかしデュダエフ政権部隊は頑強に抵抗, 双方の死者は数百人に達する見込みだ。

→ $t:DCT$ vs $t:\sim DCT, span:part$

次の例では、アノテータのタグが $t:DCT$ と $t:\sim DCT, span:P1D$ で揺れた。前者は新聞の速報性から、その日のことを記事にしたと解釈したのに対し、後者はそうとは限らないと解釈したと考えられる。

(33) 外相は、「非民営化・再国営化」の基本方針を打ち出した。

→ $t:DCT$ vs $t:\sim DCT, span:P1D$

こうした新聞ならではの書き方やテーマ、性質が、解釈をより難しくしていると考えられる。

このように、テキストに明示的に書かれていないことに対する解釈は読者によって少しずつ異なり、しかもいずれも誤りとは言えない。人がどのようにテキスト理解をしているのかをデータとして表現する方法の1つが、本研究のように、複数のアノテータの解釈を反映させたタグを全て記載することだと考える。

6 時間情報推定

本研究で付与したタグを推定する簡単なモデルを作成し、そのエラー分析を通して、推定に必要な知識や技術について議論する。

6.1 問題設定

本稿で提案した時間タグは多様な時間情報を扱える一方で、作成したコーパスの分量が大きいいため、機械学習を用いて直接推定するにはタグがスパースだという問題がある。そこでタグを下記の3つの要素へと簡略化し、各々を多クラス分類問題として推定する(図2)。

- a. 時間性：事象が時間性をもつか否か (2クラス)
- b. 事象の時間的長さ：事象の発生期間 (4クラス)
- c. 事象の発生時期：文書作成日を基準とした事象の発生時期 (5クラス)

以下に各タスクの詳細を述べる。

図 2 推定を行う 3 つのタスク

6.1.1 (a) 時間性判定タスク

対象表現が時間性をもつか否か、すなわち対応する時間タグが $t:n/a$ か否かを判定する。全対象表現 4,534 個のうち、76%の 3,438 個が時間性をもつ。

6.1.2 (b) 事象の時間的長さ分類タスク

時間性をもつ事象において、事象の時間的な長さを時間粒度に応じて 4 つのクラスに大別し、4 クラス分類問題として考える。具体的には、期間を 1 日以内、1 ヶ月未満、1 年未満、1 年以上、の 4 つに分類する。時間共参照や span:part が付与された表現など時間の長さが分からないものはデータから取り除き、全 4,534 対象表現のうち 2,752 個を用いた。以下に例を示す。

- 1 日以内: $t:1994-12-31, t:1994, \text{span:P1D}$
- 1 ヶ月未満: $t:1994-12-25 \sim 1994-12-31, t:1994-12, \text{span:P3D}$
- 1 年未満: $t:1994-12$
- 1 年以上: $t:1994, t:PAST$

全体に占めるクラスの割合は、1 日以内 (1,545 個, 56%), 1 ヶ月以内 (640 個, 23%), 1 年以上 (135 個, 5%), 1 年以上 (432 個, 16%) である。

6.1.3 (c) 事象の発生時期分類タスク

時間性をもつ事象における、文書作成日から事象を代表する日にち（以降、事象代表日と呼ぶ）までの日数を時間の粒度に応じて 5 つのクラスに大別し、5 クラス分類問題として考える。ここで、事象代表日とは事象の開始日と終了日の中間に位置する日を指す。 span タグを用いて表されるような開始日や終了日が不明な事象の場合、より広い範囲を表す t タグの中間に位置する日を用いる。例えば $t:1994-12, \text{span:P3D}$ では、1994 年 12 月の中間の日である 1994 年 12 月 16 日を事象代表日とする。

具体的には、3 年以上前、3 年前から 3 日前、3 日前から 3 日後、3 日後から 3 年後、3 年以上後の 5 つである (図 3)。時間共参照など事象の発生時期が明確に分からないものはデータから取り除き、全 4,534 対象表現のうち 3,276 個を用いた。

図 3 事象の発生時期の 5 クラス分類 (タスク c)

以下に文書作成日を 1995 年 1 月 1 日とした場合の例を示す.

- 3 年以上前: $t:1990$, $t:PAST$
- 3 年前から 3 日前: $t:1994-06-01$, $t:1994, span:P1D$
- 3 日前から 3 日後: $t:DCT$, $t:1994-12-29 \sim 1994-12-31$
- 3 日後から 3 年後: $t:1995-03, span:part$, $t:FUTURE-M$
- 3 年以上後: $t:2000$, $t:FUTURE$

全体に占めるクラスの割合は, 3 年以上前 (284 個, 9%), 3 年前から 3 日前 (879 個, 27%), 3 日前から 3 日後 (1,331 個, 41%), 3 日後から 3 年後 (571 個, 17%), 3 年以上後 (211 個, 6%) である.

6.2 モデル

事象が時間性をもつかどうかは, 対象表現中の語彙情報が大きな手がかりとなると考えられる. 例えば「出勤する」は多くの場合時間性をもつのにに対し, 「装甲」などのサ変名詞は時間性をもたないことが多い. 一方で, 事象の時間的長さや発生時期は文脈に大きく依存すると考えられる. 例えば, 対象表現「滞在した」の前に「3 日間」という表現があるのか「1 年間」という表現があるのかで事象の時間的長さは全く異なる. 同様に, 対象表現「昇段した」の前の表現が「昨日」なのか「昨年」なのかで発生時期は異なる. そこで本研究では 2 つのモデルを用意し, 時間性判定タスクでは対象表現に着目したモデルを, 事象の時間的長さ分類タスクと発生時期分類タスクでは文脈にも着目したモデルを用いる. コーパス中の対象表現が多様でスパースであること, また対象表現自体が重要な手がかりであり word2vec で事前学習した単語分散表現が有効であると考え, これを活かしたニューラルネットワークモデルを構築する. モデルの全体図を図 4 に示す.

両モデルの違いは対象表現を表すベクトルである. 時間性判定のモデルでは, 対象表現中の各単語ベクトルを足し合わせたものを用いる. 例えば, 図 4 の例 (左) では, 対象表現「空爆を」を構成する「空爆」と「を」の 2 つの単語ベクトルを足し合わせる. 文脈を用いるモデルでは, まず文全体に対して双方向 GRU (Gated Recurrent Unit) (Chung, Gulcehre, Cho, and Bengio

図 4 対象表現「空爆を」の時間情報を推定する 2つのニューラルネットワークモデル。対象表現中の語彙情報のみを用いる (a) 時間性判定モデル (左) と、文脈を使用する (b,c) 事象の時間的長さ・発生時期推定モデル (右)。

2014) を適用した後、対象表現中の自立語の単語ベクトルを用いる。図 4 の例 (右) では、対象表現「空爆を」の中の自立語である「空爆」の単語ベクトルを用いる。両モデルとも、この対象表現ベクトルに、同一文／直近の時間表現情報を表すベクトルと、時間表現との共起スコアベクトルを結合し、パーセプトロンを用いてクラス分類を行う。以下に両ベクトルの詳細を述べる。

- 同一文・直近の時間表現情報ベクトル：

文に含まれる時間表現の時間的粒度を表す 4 次元のバイナリベクトルである。文中の時間表現をルールベースで検出し、検出された時間表現が (日, 週, 月, 年) の各粒度に当てはまるか否かの 2 クラスでベクトルを構成する。例えば図 4 の例の場合、文中には「昨日」という時間表現が存在するため、「日」の粒度に該当する次元のみ 1 でその他は 0 の 4 次元ベクトル (1,0,0,0) となる。「同一文中の時間表現情報ベクトル」は、対象表現の文の時間情報ベクトルであり、「直近の時間表現情報ベクトル」は、対象文以前で時間表現を含む文の時間表現情報ベクトルである。全 4,534 対象表現のうち、35%にあたる 1,601 表現で同一文中に時間表現が検出され、また 87%にあたる 3,940 表現ではその文以前に時間表現が検出された。

- 共起スコアベクトル：

対象表現と時間表現粒度の共起度を表す 4 次元の実数値ベクトルである。まず前処理として、前項と同様にテキスト中の時間表現をルールベースで検出し、これを (日, 週, 月, 年) のいずれかの粒度に変換する。その後、対象表現と各粒度との同一文中での共起スコアを算出する。共起スコアには自己相互情報量 (Pointwise Mutual Information, PMI) を、データには 1984 年から 2005 年までの朝日新聞 1,300 万文を用いた。

6.3 実験

各タスクにおけるクラス分類には、コーパス構築時において第2段階でアノテーションされたタグを利用し、5分割交差検定を用いて学習・評価を行った。2クラス分類である時間性判定タスクの評価にはF1値、多クラス分類である事象の時間的長さ分類・事象発生時期分類タスクの評価にはMicro-F1値を用いた。

パーセプトロンには、50次元の1つの隠れ層を持つ2層フィードフォワードニューラルネットワークを使用した。損失関数には交差エントロピー誤差関数を、パラメータの最適化にはAdadeltaを使用した。GRUの隠れ層は100次元である。

単語の分散表現には98億文のWebテキストにより事前学習された200次元のベクトルを、品詞の分散表現にはランダムに初期化された10次元のベクトルを用いた。これらは誤差逆伝播時に値が更新される。

各タスクの実験結果を表8に示す。いずれのタスクも各クラスのデータ量に大きな偏りがあることから、マジョリティのクラスのみを出力した場合のスコアをベースラインとして記載した。時間性判定タスクでは約9割のスコアが得られた一方、事象の時間的長さ分類タスクは約6割、事象発生時期分類タスクは約5割にとどまった。

また、各タスクで最もスコアが良かった条件におけるConfusion Matrixを表9に示す。時間性判定タスクでは、誤りの約7割が時間性のないものをあると判定したものであった。事象の時間的長さ分類タスクでは、「1ヶ月未満」クラスを「1日以内」クラスと誤答することが多かった。事象の発生時期分類タスクでは、「3年前から3日前」クラスと「3日前から3日後」クラスの区別で多くの誤りが存在する。

表8 (a) 時間性判定タスク, (b) 事象の時間的長さ分類タスク, (c) 事象の発生時期分類タスクにおける実験結果

	時間性	事象の時間的長さ	事象の発生時期
ベースライン (マジョリティクラスを出力)	86.25	58.26	40.63
ニューラルネットワークモデル			
単語	90.09	58.76	45.57
単語+品詞	90.37	59.34	48.32
単語+品詞+共起	90.52	60.61	48.72
単語+品詞+同一文時間表現+直近時間表現	89.87	59.41	48.81
単語+品詞+共起+同一文時間表現+直近時間表現	90.35	59.81	49.69

2クラス分類である時間性判定タスクではF1値を用い、他の2つのタスクではMicro-F1値を用いて評価した。

表 9 (a) 時間性判定タスク, (b) 事象の時間的長さ分類タスク, (c) 事象の発生時期分類タスクにおける Confusion Matrix

(a) 時間性判定タスク

正解\予測	なし	あり
なし	608	488
あり	192	3,246

(b) 事象の時間的長さ分類タスク

正解\予測	1 日以内	1 ヶ月未満	1 年未満	1 年以上
1 日以内	1,266	149	15	115
1 ヶ月未満	332	178	16	114
1 年未満	35	40	14	46
1 年以上	135	74	13	210

(c) 事象の発生時期分類タスク

正解\予測	3 年以上前	3 年前から 3 日前	3 日前から 3 日後	3 日後から 3 年後	3 年以上後
3 年以上前	105	56	69	41	13
3 年前から 3 日前	63	387	333	64	32
3 日前から 3 日後	59	334	820	76	42
3 日後から 3 年後	43	112	118	237	61
3 年以上後	5	35	51	41	79

6.4 議論

6.4.1 時間性判定タスク

本タスクでは、単語分散表現の情報のみでベースラインを 3 ポイント以上上回った。これは、対象表現自体が時間性の有無に大きく関係していること示唆している。また、品詞分散表現を利用することでスコアの向上が見られた。「対して」や「総当たり」などのサ変動詞／名詞からなる時間性をもたない対象表現検出に役立ったと考えられる。

本モデルでは対象表現のみに着目したが、このアプローチでは次の例の「開発」を正しく解くことができない。

(34) 国連の支援で、総合**開発**の立案などの成果をあげた（時間性あり）

(35) **開発**のゆがみを知る人たち（(時間性なし)

(36) 経済**開発**区（時間性なし）

例 (34) は特定の開発事業であるため時間性をもつが、例 (35) や例 (36) は不特定あるいは一般的な事象であるため時間性をもたない。このような例に対処するためには、対象表現の前後の表現や、対応する述語や項の性質などの情報を考慮する必要がある。

6.4.2 事象の時間的長さ分類タスク

本タスクでは共起スコアベクトルの導入によりスコアが向上した。例えば次の例の「宿泊」は、単語と品詞情報のみのモデルでは誤答したが、「宿泊」と日にち粒度の時間表現の共起スコ

アが高いという情報を与えることで正答した.

(37) ホテルの宿泊者が目撃した

一方, 表 9 から分かるように, 多くの例で誤って「1 日以内」クラスを出力した. 次の例は全て「1 日以内」クラスを出力し誤答したものである.

(38) 兵士が首都から南に脱出している

(39) 見逃せないのは労組の圧力だ

(40) 出稼ぎ世帯の大半は、テレビやバイクを買う

(41) 啓蒙に力を入れている

例 (38), 例 (39), 例 (40) はいずれも, 対象表現自体は 1 日以内の事象とも捉えることができるものであるが, この文脈ではそうではない. 本モデルは文脈を考慮するものであるが, より大規模なデータで学習する必要があると考えられる. 例 (41) では, 対象表現は軽動詞であり, 「啓蒙」が時間を考える上での手がかりである. このような単語の含意する時間情報知識を大規模に獲得し, モデルに取り入れる必要がある.

6.4.3 事象の発生時期分類タスク

本タスクでは, 時間表現情報ベクトルがスコア向上に貢献している. これは例えば「四日発生した雪崩」のように, 時間表現が重要な手がかりを与える場合が多く存在するためだと考えられる. 本モデルでは「昨日」や「1995 年」などの明示的な時間表現のみを対象としたが, テキスト中には「ベトナム戦争」や「前回のワールドカップ」など暗黙的に時間情報を表す表現も多く存在し, 誤答の原因となっている. Wikipedia などの外部知識を用いて, より広範な時間情報を考慮する必要がある.

本モデルでは時間表現情報を 4 つの粒度で表したが, テキスト中には「昼過ぎ」などより細かい粒度の情報も多く存在する. これらの情報を利用することでスコア向上が期待される. また, 本モデルでは時間の粒度のみに着目したが, 時間表現が日付を表すのか期間を表すのかで大きく意味が変わる. 例えば「3 日」という時間表現を「1 月 3 日」のことと解釈するのか「3 日間」のことと解釈するのかは, 事象の時間的長さ・発生時期の推定タスクにおいて非常に重要であり, さらなるスコア向上のためには高度な時間表現解析器の導入が必要である.

7 本アノテーションの応用

本アノテーションには様々な応用先が考えられる. 例えば, ニュース記事からのタイムラインやサマリーの生成である. ニュースには事件や出来事に関する新しい情報は記載されるが,

それを取り巻く出来事やこれまでの経緯は記載されないことも多いため、出来事の全体像を知るためには複数の記事を読み情報を統合する必要がある。またニュースには、その時の情報に基づいた予想や解釈も記載されるため、異なる時期に書かれた別の記事と比較することは情報の信頼性を担保する上で有益である。

本稿で述べたアノテーション基準やデータ、時間情報推定モデルは、このような記載時期の異なる情報の統合や比較に利用することができる。図5は、1995年1月に行われたサッカー大会に関する事象を時間軸上に配置したものである。文章作成日を区別して事象を配置するため2つの時間軸を導入しており、横軸は文章作成日、縦軸は事象の起きた日を表す。例えば右上の「デンマークが優勝した」は、1月15日に記述された事象で、これに対応する時間値は1月13日 ($t:1995-01-13$) である。本研究で作成したデータは、このようにアプリケーションを構築するためのモデルの学習や評価に用いることができる。本研究でアノテーションを行った京都大学テキストコーパスには、既に述語項構造や共参照関係のアノテーションがなされているため、「ラモス」など特定のエンティティに着目した学習や評価にも用いることが可能である。

また、本アノテーション基準や時間情報推定モデルを利用することにより、事象に関する時間情報知識を大量のテキストから収集することが可能となる。出来事や状態の発生日や継続日数、頻度などの情報は、情報検索、要約、対話、文生成など多くのアプリケーションで利用す

図5 1995年1月に行われた、サッカーのインタコンチネンタル選手権に関する事象の時間軸への配置。横軸は文章作成日、縦軸は事象の起きた日を表す。例えば右上の「デンマークが優勝した」は文章作成日が1月15日で、これに対応する時間値は1月13日である。

ることができる。本稿で述べたアノテーション基準は、これらの時間情報を詳細に記述する手段として用いることができ、また時間情報推定モデルは事象に関する時間情報をテキストから自動獲得する技術に応用することが可能である。

8 おわりに

本稿では、テキスト中の時間情報を網羅的にアノテーションするための新しいタグ付け基準について述べた。従来研究よりも広範な表現をアノテーション対象とし、また表現力豊かな時間タグを導入した。京都大学テキストコーパス中の 113 文書に対してアノテーションを行った結果、76%の表現が時間性をもつと判断され、そのうちの 35% (全体の 26%) に本稿で新たに提案された時間タグが付与された。本研究で付与した時間情報を、時間性・事象の時間的長さ・事象の発生時期の 3 つの観点から推定するモデルを構築した。各精度は F 値において 90%, 61%, 49%であった。

本研究では、新聞記事を対象にアノテーションを行ったが、新聞特有の書き方や性質がアノテーションの揺れの原因の 1 つになっていると考えられる。今後は、Web など新聞以外のコーパスでのアノテーションを試みたい。

また、本研究では時間表現に対するアノテーションは行っていない。これは、日本語時間表現に対するアノテーションにおけるアノテータ間一致度は高いことが (小西, 浅原, 前川 2013) により報告されており、時間表現に対する解釈が揺れて事象に対するアノテーションに影響を及ぼすことは少ないと考えたからである。しかし、時間情報解析器を構築する上では有益な情報になると考えられるため、今後は時間表現にもアノテーションを行い、より充実した時間情報コーパスにしていきたい。

謝 辞

本研究の一部は JST CREST JPMJCR1301, AIP チャレンジの助成によるものです。また本論文の内容の一部は、11th edition of the Language Resources and Evaluation Conference で発表したものです (Sakaguchi, Kawahara, and Kurohashi 2018)。

アノテーションにご協力いただいた石川真奈見氏, 堀内マリ香氏, 二階堂奈月氏に感謝いたします。また、時間情報推定モデルに関して有益なコメントをくださった柴田知秀氏に感謝いたします。

参考文献

- Asahara, M., Kato, S., Konishi, H., Imada, M., and Maekawa, K. (2014). “BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text.” *International Journal of Computational Linguistics & Chinese Language Processing*, **19** (3), pp. 1–24.
- Asakura, Y., Hangyo, M., and Komachi, M. (2016). “Disaster Analysis using User-Generated Weather Report.” In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, pp. 24–32, Osaka, Japan. The COLING 2016 Organizing Committee.
- Cassidy, T., McDowell, B., Chambers, N., and Bethard, S. (2014). “An Annotation Framework for Dense Event Ordering.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 501–506, Baltimore, Maryland. Association for Computational Linguistics.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*.
- Hayes, A. F. and Krippendorff, K. (2007). “Answering the Call for a Standard Reliability Measure for Coding Data.” *Communication Methods and Measures*, **1** (1), pp. 77–89.
- Huang, R., Cases, I., Jurafsky, D., Condoravdi, C., and Riloff, E. (2016). “Distinguishing Past, On-going, and Future Events: The EventStatus Corpus.” In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 44–54. Association for Computational Linguistics.
- Kawahara, D., Kurohashi, S., and Hasida, K. (2002). “Construction of a Japanese Relevance-tagged Corpus.” In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands—Spain. European Language Resources Association (ELRA).
- Kolomiyets, O., Bethard, S., and Moens, M.-F. (2012). “Extracting Narrative Timelines as Temporal Dependency Structures.” In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Republic of Korea, 8-14 July 2012*, pp. 88–97. ACL.
- 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. 201–221.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology (second edition)*. Sage Publications.
- Minard, A.-L., Speranza, M., Agirre, E., Aldabe, I., van Erp, M., Magnini, B., Rigau, G., and Urizar, R. (2015). “SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering.” In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*,

- pp. 778–786, Denver, Colorado. Association for Computational Linguistics.
- Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Beth Sundheim, D. D., Ferro, L., and Lazo, M. (2003). “The TIMEBANK Corpus.” In *Proceeding Corpus Linguistics 2003*, pp. 647–656.
- Reimers, N., Dehghani, N., and Gurevych, I. (2016). “Temporal Anchoring of Events for the TimeBank Corpus.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2195–2204, Berlin, Germany. Association for Computational Linguistics.
- Sakaguchi, T., Kawahara, D., and Kurohashi, S. (2018). “Comprehensive Annotation of Various Types of Temporal Information on the Time Axis.” In *Proceedings of the 11th Edition of its Language Resources and Evaluation Conference*, pp. 332–338, Miyazaki, Japan.
- Sauri, R., Littman, J., Gaizauskas, R., Setzer, A., and Pustejovsky, J. (2006). “TimeML Annotation Guidelines, Version 1.2.1.”
- 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 Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 1–9. Association for Computational Linguistics.
- Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, 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. Association for Computational Linguistics.
- Verhagen, M., Sauri, R., Caselli, T., and Pustejovsky, J. (2010). “SemEval-2010 Task 13: TempEval-2.” In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62. Association for Computational Linguistics.

略歴

坂口 智洋：2013年東京大学理学部生物情報科学科卒業。2015年京都大学大学院情報学研究科修士課程修了。同大学院博士課程在学中。修士（情報学）。自然言語処理の研究に従事。

河原 大輔：1997年京都大学工学部電気工学第二学科卒業。1999年同大学院修士課程修了。2002年同大学院博士課程単位取得認定退学。東京大学大学院情報理工学系研究科学術研究支援員、独立行政法人情報通信研究機構研究員、同主任研究員を経て、2010年より京都大学大学院情報学研究科准教授。自然言

語処理, 知識処理の研究に従事. 博士 (情報学). 情報処理学会, 言語処理学会, 人工知能学会, 電子情報通信学会, ACL, 各会員.

黒橋 禎夫: 1994 年京都大学大学院工学研究科電気工学第二専攻博士課程修了. 博士 (工学). 2006 年 4 月より京都大学大学院情報学研究科教授. 自然言語処理, 知識情報処理の研究に従事. 言語処理学会 10 周年記念論文賞, 同 20 周年記念論文賞, 第 8 回船井情報科学振興賞, 2009 IBM Faculty Award 等を受賞. 2014 年より日本学術会議連携会員.

(2018 年 8 月 1 日 受付)

(2018 年 11 月 1 日 再受付)

(2018 年 12 月 12 日 採録)