

京都大学テキストコーパスに対する 網羅的な時間情報アノテーション

坂口 智洋^{1,a)} 河原 大輔^{1,b)} 黒橋 禎夫^{1,2,c)}

概要: テキストの時間的意味理解を目的として、事象の時間的順序関係解析やタイムライン生成などの研究が盛んに行われている。これらの研究においてモデルの学習や評価を行うために、事象情報と時間情報を関連付けたコーパスが開発されてきた。本研究では、テキスト中の様々な表現がもつ時間情報を網羅的かつ表現力豊かにアノテーションすることを目的として、先行研究を次の2点で拡張したアノテーション基準を提案する。1つは従来扱われてきた一時性の強い事象表現だけでなく、状態のような一時性は弱いものの時間性をもつ表現も対象に含めること、もう1つは言語表現がもつ多様な時間情報を扱えるよう、より表現力の高い時間タグを導入することである。このアノテーション基準にもとづいて、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。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である。

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り、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 [13], [14], [15] 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、また SemEval 15 [9] では複数のテキストから事象表現を抽出しこれを時系列に配置するタイムライン生成タスクが扱われた。

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 [3], [10], [11]。これらのコーパスでは一時性の強い事象表現を対象としたアノテーションが行われたが、人々がテキスト中の時間情報をどのよう

に理解しているかを知るためには、時間性の曖昧な表現も含めてどのように解釈されているかを正確に知る必要がある。本研究では、このような表現を含めた、テキスト中の様々な表現がもつ時間情報を網羅的かつ表現力豊かにアノテーションする。

本アノテーション基準の特徴は次の2つである。1つは、時間性をも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 [12] のガイドラインで定義された、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“event”ではない表現がもつ時間情報もテキスト解釈の手がかりとなり得る。上の例の場合、「出現しており」が数年前から現在にかけての事柄であるという時間情報は、テキストの含意する常識を明確化するためにタグ付けする必要があると考える。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もち得る全ての事象表現、すなわちテキスト中の全ての述語と事態性名詞(以降、対象表現と呼ぶ)をアノテーション対象とする。

我々の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時間軸に対応付けられることである。[11]は事象の起きる期間を開始点と終了点を用い

¹ 京都大学 大学院情報学研究所

² 科学技術振興機構 CREST

a) sakaguchi@nlp.ist.i.kyoto-u.ac.jp

b) dk@i.kyoto-u.ac.jp

c) kuro@i.kyoto-u.ac.jp

て表したが、次の例のような「飛び飛びな時間」や「大きな区間の中のある一部の期間」などの表現を正確に時間軸に結びつ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の表現が表す多様な時間情報を、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。

上記のような多様な表現に対して表現力豊かなタグを付与することにより、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、対象表現の約80%に時間性が認められ、これらのうち約40%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である。

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、これまでも多く作られており、これらは大きく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 [10]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 [1]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。[7]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 [3]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 [5]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。[2]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。[11]は、TimeBank Corpus中の全て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日以内で終わるものか、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日で終わる事象であり、時間値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」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**、米国人女性初の宇宙遊泳を行った。2014年にタイム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日で終わる事象、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。

本研究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

本研究では、テキスト中の全ての述語と事態性名詞に対して時間タグを付与する。まずテキストに対して形態素

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>
	TBU の組み合わせ で表される時間値	2. 漠然とした時間タグ (過去・現在・未来)	(例) <i>t:PRESENT</i>
		3. 相対的な時間タグ (時間共参照・発話日)	(例) <i>t:選挙</i>
時間性なし		a. TBU 間の期間 (TBU~TBU)	(例) <i>t:1995-01-05~1995-01-07</i>
		b. TBU 中の特定の幅をもつ期間 (span)	(例) <i>t:1995-01,span:P1W</i>
		c. TBU 中の不特定の期間 (span:part)	(例) <i>t:1995-01,span:part</i>
		d. TBU 中の繰り返し (freq)	(例) <i>t:1995-01,freq:2/P1W</i>
			(例) <i>t:n/a</i>

表 1 時間タグの一覧

解析を行い、述語と事態性名詞の基本句を抽出する。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時 (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 *not applicable* (*t:n/a*) 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、特定の時点を表すもので、3種類のタグを定義する。また、これらを組み合わせることで期間を表現する。組み合わせ方は4通りあり、多様な時間情報を表現できる。[11] で扱った時間タグは、表 1 中の時間基本単位 1 と組み合わせ a と c のみである。

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 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 日のこととして考える。

(7) [DCT: 2017-04-29] 今日は夜 6 時に帰った。

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どうかに着目して判断する。

時間性をもつ例をいくつか挙げる。

(8) 明日京都に行く。

(9) 言語処理研究が盛んだ。

(10) あの子は背が高い。

例 (8) 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 (9) 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のため時間性をもつと考える。例 (10) の「高い」も、ずっと背が高かったわけではないため、時間性をもつと考える。先行研究がアノテーション対象としているのは、例 (8) のみである。

次に時間性をもたない例を示す。

(11) ウサギは草を食べる。

(12) 彼の目は黒い。

例 (11) の「食べる」は、昔から変わらない事柄であるため時間性をもたないと考える。例 (12) 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。

(13) 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」には、時間値 *t:2017-04-28* を付与する。

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

→ *t:2017-04-08*

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、8 月 1 日から 31 日までちょうど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。

(15) [DCT: 2017-04-29] 昨年の 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。例えば次の例の「行く」の時間値は *t:DCT+P1W* と表せる。

(16) 私も来週そこに行く。

→ *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を表しているのかわからない。

(17) 昔広島に住んでいた。

[11] 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* とタグ付けしたが、本研究ではこのような漠然とした言語表現に対応するタグを導入する。

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」は今日より少し前から少し後までという意味を含む。例えば次の文の「持ち込める」は、今日だけでなく、今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。

(18) 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。「80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。

(19) 1980年頃に**建てられた**建物

→ *t:1980ap*

3.2.3 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点である。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、絶対的な時間値がわ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わからないが、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、具体的な日付はわからないが「選挙」の翌日であることは分かる。この場合、*t:選挙+PID* と記す。

(20) 選挙の翌日、大規模なデモが**起きた**。

*1 数週間前までの過去は、3.3.1節の～を用いて *t:~DCT* と表す。

→ *t:選挙+PID*

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。

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、わ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~* を付与する。

(21) 「これから**頑張るしかない**」と監督は言った。

→ *t:UD~*

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) の組み合わせ

3.3.1 TBU 間の期間を表す方法

開始時を表す TBU と終了時を表す TBU を～で結ぶことで TBU と TBU の間の期間の時間値を表す。これは [11] 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*

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:~2017-04-28* となる。

(22) [DCT: 2017-04-29] 昨日まで**忙しかった**。

→ *t:~2017-04-28*

遠い過去や未来である場合は、*PAST* や *FUTURE* を用いる。

3.3.2 TBU 内の一部の期間を表す方法

ある大きな TBU の中の一部の期間を表す場合、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。期間の長さがわからない場合は *span:part* と記す。例 (6) の場合、「滞在し」は1984年10月のどこかなので *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:PID* 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**、米国人女性初の宇宙遊泳を行った。2014年にタイム100に**選ばれた**。

	アノテータ 1	アノテータ 2	アノテータ 3	平均
日付タグ	1101 (24.3%)	1062 (23.4%)	918 (20.3%)	1027 (22.7%)
<i>t:DCT</i>	276 (6.1%)	256 (5.7%)	159 (3.5%)	230 (5.1%)
漠然とした時間を含むタグ	964 (21.3%)	825 (18.2%)	797 (17.6%)	862 (19.0%)
<i>PRESENT</i> を含むタグ	586 (12.9%)	286 (6.3%)	219 (4.8%)	364 (8.0%)
<i>PAST</i> を含むタグ	134 (3.0%)	181 (4.0%)	255 (5.6%)	190 (4.2%)
<i>FUTURE</i> を含むタグ	286 (6.3%)	363 (8.0%)	324 (7.2%)	324 (7.2%)
<i>ap</i> を含むタグ	6 (0.1%)	16 (0.4%)	16 (0.4%)	13 (0.3%)
相対的な時間を含むタグ	351 (7.7%)	314 (6.9%)	565 (12.5%)	410 (8.0%)
時間共参照を含むタグ	138 (3.0%)	58 (1.3%)	207 (4.6%)	134 (3.0%)
<i>UD</i> を含むタグ	212 (4.7%)	259 (5.7%)	366 (8.1%)	279 (6.2%)
TBU 間の期間 (~)	1377 (30.4%)	1528 (33.7%)	1903 (42.0%)	1603 (35.3%)
TBU 中の特定の幅をもつ期間 (span)	585 (12.9%)	477 (10.5%)	624 (13.8%)	562 (12.4%)
TBU 中の不特定の期間 (span:part)	499 (11.0%)	617 (13.6%)	527 (11.6%)	548 (12.1%)
TBU 中の繰り返し (freq)	47 (1.0%)	52 (1.2%)	49 (1.1%)	49 (1.1%)
時間性をもたないタグ (<i>t:n/a</i>)	1071 (23.6%)	1078 (23.8%)	1056 (23.3%)	1068 (23.5%)
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%)
本稿で新たに提案されたタグ	1512 (33.4%)	1183 (26.1%)	1407 (31.0%)	1367 (30.2%)
合計	4534	4534	4534	4534

表 2 アノテーションされた時間タグの分布

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

[11] 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。

3.3.3 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」に行う事柄や「3日に1回」行う事柄もある。このように飛び飛びで複数日に渡って起きる表現は、*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。

繰り返しに関する時間タグは次の3つの記述方法がある。

- 「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* とタグ付ける。

(23) [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。

(24) 天神市は毎月25日に開催される。

→ *t:PRESENT,freq:@@@@-@@-25*

(25) 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わからない場合、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。次の例の場合、「行く」には *t:PRESENT,freq:sometimes* を付与する。

(26) [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 [6] 中の113文書856文、4,534表現に対して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、本研究で時間タグを付与することで複数のテキストからエンティティに着目したタイムライン生成や時系列に着目した情報検索など、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である。

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つをローテーションで担当する。これにより、最終的に全員が全ての文書に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時間タグの分布を表2に示す。約2割の対象表現が時間性をもたないと判定された。日付タグと漠然とした時間タグがそれぞれ全体の約2割を占める一方で、相対的なタグは約1割と少ない。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3割を占めた。

4.4 アノテータ間一致率

Krippendorff's α [4], [8] を用いてアノテータ間一致率を算出した。[11]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。もう1つは時間タグの部分一致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:~1994-12-31* では、両者の範囲が重なっているため一致と判定する。一方、*t:~1994-12-31* と *t:1995-01-01* のように、一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。各段階での一致率を表3に記す。最終段階とは、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、最終段階ともに約1,300個)を除いたものである。

第1段階と最終段階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、前者は独立にタグ付けを行ったのに対し、後者では第2段階で他のアノテータのタグを見ることができた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、Relax 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。

[11]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。

	Strict 基準	Relax 基準
第1段階	0.417	0.719
最終段階	0.554	0.802
第1段階 (<i>t:n/a</i> を除く)	0.380	0.803
最終段階 (<i>t:n/a</i> を除く)	0.526	0.867
[Reimers+ 16]	0.617	0.912

表3 Krippendorff's α を用いて算出したアノテータ間一致率

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るかを分析する。

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4、表5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。表4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5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ことから、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領域が重なっており、この基準では一致したことが分かる。

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。

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である。*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とするか、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っている。

次の例では、アノテータのタグが *t:PRESENT* と *t:n/a*

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	PRES	n/a	110
DAY	DAY	740	DAY	n/a	105
~DAY,span	~DAY,span	142	DAY	~DAY,span	104
PRES	PRES	113	~DAY,span	~DAY,span	77
DAY~,span	DAY~,span	54	DAY	~DAY	73
DAY~DAY	DAY~DAY	38	PRES	~DAY,span	59
YEAR	YEAR	12	PRES	~DAY	49
DAY~FUTURE	DAY~FUTURE	10	PRES	PAST~DAY	49
合計		2045	合計		2275

表 4 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
n/a	n/a	800	PRES	n/a	110
DAY	DAY	741	DAY	n/a	105
~DAY,span	~DAY,span	219	DAY	DAY	44
PRES	PRES	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	PRES	59	DAY~DAY	n/a	14
合計		3533	合計		787

表 5 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

で揺れた。

(27) 大統領官邸のある中心部

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

[11] も指摘しているとおり、ある事象がある日一日で終わることなのか、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 の揺れとして表れる。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、記事執筆日に起きたのか、執筆日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。前者はこの表現を執筆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。

(28) 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。

次の例では、アノテータのタグが $t:DCT$ と $t:\sim DCT,span:PID$ 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。

(29) 外相は、「非民営化・再国営化」の基本方針を**打ち出した**。

このような、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。

6. 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。

個人の文章解釈や常識を反映させるため、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、約 8 割の表現が時間性をもつと判断され、そのうちの約 4 割に本稿で新たに提案された時間タグが付与された。アノテータが独立にタグ付けを行ったときのアノテータ間一致率は、Krippendorff's α において Strict 基準では 0.417、Relax 基準では 0.719 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。

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。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。

参考文献

- [1] Asahara, M., Kato, S., Konishi, H., Imada, M. and Maekawa, K.: BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text, *International Journal of Computational Linguistics & Chinese Language Processing, Volume 19, Number 3, September 2014*, (online), available from <http://aclweb.org/anthology/O14-4001> (2014).
- [2] Asakura, Y., Hangyo, M. and Komachi, M.: Disaster Analysis using User-Generated Weather Report, *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka, Japan, The COLING 2016 Organizing Committee, pp. 24–32 (online), available from <http://aclweb.org/anthology/W16-3906> (2016).
- [3] Cassidy, T., McDowell, B., Chambers, N. and Bethard, S.: An Annotation Framework for Dense Event Ordering, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimore, Maryland, Association for Computational Linguistics, pp. 501–506 (online), available from <http://www.aclweb.org/anthology/P14-2082> (2014).
- [4] Hayes, A. F. and Krippendorff, K.: Answering the Call for a Standard Reliability Measure for Coding Data, *Communication Methods and Measures*, Vol. 1, No. 1, pp. 77–89 (online), DOI: 10.1080/19312450709336664 (2007).
- [5] Huang, R., Cases, I., Jurafsky, D., Condoravdi, C. and Riloff, E.: Distinguishing Past, On-going, and Future Events: The EventStatus Corpus, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 44–54 (online), available from <http://aclweb.org/anthology/D16-1005> (2016).
- [6] Kawahara, D., Kurohashi, S. and Hasida, K.: Construction of a Japanese Relevance-tagged Corpus., *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands - Spain, European Language Resources Association (ELRA), (online), available from <http://www.lrec-conf.org/proceedings/lrec2002/pdf/302.pdf> (2002).
- [7] Kolomyiets, O., Bethard, S. and Moens, M.-F.: Extracting narrative timelines as temporal dependency structures, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, The 50th annual meeting of the association for computational linguistics, Jeju, Republic of Korea, 8-14 July 2012*, ACL, pp. 88–97 (online), available from <https://lirias.kuleuven.be/handle/123456789/347253> (2012).
- [8] Krippendorff, K.: *Content Analysis: An Introduction to Its Methodology (second edition)*, Sage Publications (2004).
- [9] Minard, A.-L., Speranza, M., Agirre, E., Aldabe, I., van Erp, M., Magnini, B., Rigau, G. and Urizar, R.: SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering, *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, Association for Computational Linguistics, pp. 778–786 (online), available from <http://www.aclweb.org/anthology/S15-2132> (2015).
- [10] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Beth Sundheim, D. D., Ferro, L. and Lazo, M.: The TIMEBANK Corpus, *Proc. Corpus Linguistics 2003*, pp. 647–656 (2003).
- [11] Reimers, N., Deghani, N. and Gurevych, I.: Temporal Anchoring of Events for the TimeBank Corpus, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, Association for Computational Linguistics, pp. 2195–2204 (online), available from <http://www.aclweb.org/anthology/P16-1207> (2016).
- [12] Sauri, R., Littman, J., Gaizauskas, R., Setzer, A. and Pustejovsky, J.: TimeML Annotation Guidelines, Version 1.2.1 (2006).
- [13] 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, *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Association for Computational Linguistics, pp. 1–9 (online), available from <http://aclweb.org/anthology/S13-2001> (2013).
- [14] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G. and Pustejovsky, J.: SemEval-2007 Task 15: TempEval Temporal Relation Identification, *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Association for Computational Linguistics, pp. 75–80 (online), available from <http://aclweb.org/anthology/S07-1014> (2007).
- [15] Verhagen, M., Sauri, R., Caselli, T. and Pustejovsky, J.: SemEval-2010 Task 13: TempEval-2, *Proceedings of the 5th International Workshop on Semantic Evaluation*, Association for Computational Linguistics, pp. 57–62 (online), available from <http://aclweb.org/anthology/S10-1010> (2010).