

主観的な時間表現を含む 時間区間アノテーションデータセットの構築

久保 旭^{1,a)} 亀甲 博貴^{2,b)} 原島 純^{3,c)} 木村 明日香^{3,d)} 関野 樹^{4,e)} 森 信介^{2,f)}

概要: テキスト中に含まれる時間表現には「夏」や「月末」など、話者の主観に基づく表現が用いられることがあるが、こうした表現は絶対的な値を定めることが困難である。本論文では、時間表現の客観性を表現可能なアノテーション基準の設計及びアノテーションを行い、時間表現アノテーションコーパスを構築する。本アノテーション基準では、主観的な時間表現を時間幅のある絶対的な時間区間として表現し、これによって時間表現を統一的に取り扱うことを可能とする。加えて、構築したコーパスを用いた時間表現のタグ推定、絶対値推定を行い、その結果を報告する。

Construction of Time Interval Annotation Dataset including Subjective Time Expressions

1. はじめに

テキスト中の時間表現は、抽出された時間情報は情報検索、情報抽出などに用いられ、それらの抽出のために多くの研究が行われてきた。時間表現の抽出は、2種類のタスクに分けることができる。一方はある文字列が時間表現であるかどうかを推定する認識タスクであり、もう一方は抽出した時間表現を日時の値として表現する正規化タスクである。これらは総称して TERN (Temporal Entity Recognition and Normalization) と呼ばれる。

時間情報認識、時間情報抽出のいずれにおいても、ルールベース、機械学習、深層学習など、様々な推定手法が提案されているが、それらの推定器の構築は予め時間表現と値をアノテーションしたデータセットを用いて行われることが多い。

アノテーションデータの構築に当たって、現在広く用いられている基準には TimeML [1] があり、さらに、TimeML

では表現が困難な時間表現をアノテーション可能とする基準の提案もなされている。

これらのアノテーション基準は、時間情報を網羅的にアノテーションすることに注目しており、1通りに定まらない絶対値の表現は曖昧さを残したままの値として表現していた。

しかし、これらの値が具体的な数値として表現されていなければ日時を用いた情報検索や値の図示は困難である。そこで、本研究では主観的な時間表現を時間軸上にプロットすることを可能にするためのアノテーション基準を提案する。これによって、1通りに定まる時間表現と統一的な枠組みによって処理を行うことが可能となる。

本アノテーション基準の特徴は2点ある。1点目は全ての時間表現を始点と終点をもつ時間区間として取り扱うことであり、2点目は時間表現を客観的かどうか、絶対的かどうかで区別し、それぞれに対してタグを付与していることである。

本研究では、この基準を用いて新聞記事のアノテーションを行い、またこのアノテーションデータを用いて推定モデルを構築し、その精度について評価を行った。

2. 関連研究

時間表現の抽出においては、TimeML [1] において提案された TIMEX3 タグによるアノテーションが広く利用されてい

¹ 放送大学教養学部
² 京都大学学術情報メディアセンター
³ 株式会社 Linfer
⁴ 国際日本文化研究センター
^{a)} 2210087910@campus.ouj.ac.jp
^{b)} kameko@i.kyoto-u.ac.jp
^{c)} j.harashima@gmail.com
^{d)} kimura.asuka@linfer.jp
^{e)} sekino@nichibun.ac.jp
^{f)} forest@i.kyoto-u.ac.jp

る。TimeML では、時間表現を DATE (日付表現), TIME (時刻表現), DURATION (期間表現), SET (頻度集合表現) の 4 タイプに分類している。BCCWJ-Timebank [2] は日本語に特化した TimeML 準拠のデータセットである。BCCWJ-Timebank では、TimeML のアノテーション基準を現代書き言葉均衡コーパスに適用し、日本語特有の時間表現をアノテーションするために、値の表現に関する拡張を提案している。また、日本語による医療言語処理のコーパス構築 [3] に TimeML が用いられている。

一方、TimeML のアノテーション基準の課題に着目し、新たなアノテーションの枠組みの提案も行われている。文献 [4] では、例えば「Saturdays since March 6」のように TimeML ではアノテーションできない事例があることを指摘したうえで、SCATE (Semantically Compositional Annotation of Time Expressions) という新たなアノテーション基準を提案している。この基準では、時間表現の構成要素を Period と Interval に分類し、それらを演算子で操作することによって時間軸上に表現するという提案がなされている。

TimeML や SCATE はコーパス上の時間表現を網羅的にアノテーションすることを志向したものであり、曖昧性のある絶対値はそのまま記述することを許容している。例えば「2024 年夏」の絶対値は、TimeML では 2024-SU と表現され、SCATE では年と季節が分割されてそれぞれ 2024, Summer と表現されるが、これをどのように時間軸上にプロットするかは定義されていない。一方、本研究は絶対値を時間軸上にプロットすることに注目しており、これらの曖昧性を解消することによって、情報検索や可視化などへの応用可能性を拡大することを目的としている。曖昧性を解消するためには何らかの方法によってこれらの表現を時間軸上にプロットできる値に置き換える必要があるが、その結果が著者や受け手にとって常に受容可能とは限らない。また、応用によって置換先の絶対値を変更する必要がある場合も考えられる。そこで、時間表現の再分類を行い、主観に基づく曖昧性がある表現について、タグによる分離を試みる。

3. 時間表現のアノテーション

3.1 値の表現

時間表現の値は、厳密にはある 1 点としてではなく、区間として表現される。本研究においては、文献 [5] と同様に時間軸上の時間表現の値を (始点, 終点) として定義する。ここで、始点は閉区間、終点は開区間である。

始点、終点の値は、ISO8601-1:2019 に準じた形式でアノテーションする。省略されている時間粒度の値は 0 又は 1 であると見なす。また、終点が省略されている場合は、始点の最も詳細な時間粒度に 1 を加えたものを終点とする。例えば「2024-01-01」の場合、[2024-01-01T00:00:00.00...

2024-01-02T00:00:00.00...) として解釈する。

3.2 時間表現タグの設計及びアノテーション

本研究では、時間表現を客観的・主観的、絶対・相対の 2 軸で分類する。

客観的表現・主観的表現 ある種の時間表現は、その絶対値が著者によらず 1 通りに定まる。本研究では、このような表現を客観的な時間表現と呼ぶ。ISO8601-1 や JIS X 0301 が注目しているのはこのような表現である。一方、著者などによって指し示す絶対値が異なる時間表現がある。「夏」や「月末」がその例であり、これらの表現が指し示す範囲について必ずしも共通認識がもたれているわけではない。本研究では、これらを主観的な時間表現と呼ぶ。

絶対表現・相対表現 3.1 の時間区間は、曖昧性解消を行ったうえで時間軸上にプロットすることができる。一方、「前日」や「1 年後」などの表現は、それ自体を時間軸上にプロットすることはできず、時間区間をシフトさせる役割をもっている。本研究では、前者を絶対表現といい、後者を相対表現という。

なお、「毎週火曜日」などの頻度集合表現は無限に繰り返される不連続な区間であると考えられるが、本研究では有限な幅をもつ時間区間に注目していることから、アノテーションの対象外としている。

上述した観点に基づいて本研究で定義した時間表現タグを表 1 に示す。

アノテーションにおいては、各時間表現に対して表 1 の中から適切なタグを選び、スパンに対してタグ付けを行う。スパンの範囲は超短単位*1を最小単位とし、絶対表現・相対表現の中での連続表現はスパンをまとめる。タグの重複は不可としているので、T タグと TS タグの連続など、異なるタグの連続表現をまとめる場合は、まとめたスパンに対してタグの再割当を行う必要がある。

3.3 時間表現間の関係のアノテーション

BCCWJ-Timebank では、時間表現の値を次の 2 種類に分類している。

定時間情報表現 表層の情報だけで正規化できる表現

不定時間情報表現 文脈の情報を用いなければ正規化できない表現

本研究では、不定時間情報表現に対して、その表現を定時間情報に正規化するための情報をアノテーションする。具体的には、基準日時へのアンカー、基準日時から抽出する時間粒度の情報、基準日時に対して行う操作の 3 情報である。

基準日時へのアンカー 本研究では、不定時間情報表現は

*1 文献 [6] の短単位に対して活用語尾を切り出し、新たに「語尾」という品詞を与えたもの。

表 1 時間表現タグ
Table 1 Temporal expression tags.

種別	タグ	概略	時間表現の例
絶対表現 客観的	T	直接表現	2024 年 1 月 1 日
	TA	概数表現がある	2024 年 1 月頃
	TS	範囲が話者による	2024 年 1 月 1 日朝
	TSA	話者によりかつ概数表現がある	21 世紀前半頃
相対表現 客観的	RT	基準時点からの相対表現	3 日前, 今年, 5 年後
	RTA	概数表現がある	3 日ぐらい前, 5 年くらい後
	RTS	範囲が話者による	明朝
	RTSA	話者によりかつ概数表現がある	昨春頃

文書中の定時間情報表現の情報をを用いて絶対値に変換可能である、と仮定する。そこで、スパン間の関係をアノテーションすることによって基準日時を示す*2。基準日時は文書作成日時を指すこともあるが、必ずしもこれに限らない。例えば、会話文を含む記事の場合、記事の作成日時と発話の日時が異なることがある。そのような場合は、適切な基準日時を選択し、アノテーションする必要がある。

基準日時から抽出される時間粒度 不定時間情報表現において欠落した情報を定時間情報表現から補完する際、基準日時から抽出する時間粒度をアノテーションする。そのために、文献 [4] において提案されている暦の時間粒度を一部変更して抽出対象となる時間粒度のタグを定義した。タグの一覧を表 2 に示す。なお、CenturyOf は 1 年が起点であり、TwoDigitYearOf は 0 年が起点*3である。

基準日時に対して行う操作 本研究では、文献 [4], [7] において提案されている temporal operators の概念を取り入れ、前項の粒度で取り出した基準日時に対する演算操作を行うことで不定時間情報表現を定時間情報表現に変換することを想定する。演算子としては、加算 (After)、減算 (Before)、加減算を行わない演算 (This) の 3 種類を用いる。

加減算を行う際は、引数として加減算する量を与える必要がある。引数の値はデフォルトを 1 とする。相対表現の場合は、絶対値の推定処理において、エンティティから加減算の量の抽出を行い、その量によって置き換えられる。

絶対表現の場合は、任意の時間粒度で取り出された基準日時を操作した結果と不定時間情報表現とをマージすることによって、定時間情報表現を復元することができる。相対表現の場合は、基準日時を操作した結果に対し、相対表現が示す時間量を更に加減算すること

表 2 時間粒度のタグ
Table 2 Temporal granularity tags.

時間粒度	タグ	時間表現の例
世紀	CenturyOf	<u>21 世紀</u>
元号	EraOf	<u>令和 6 年 1 月 1 日</u>
元号年	EraYearOf	<u>令和 6 年 1 月 1 日</u>
年	YearOf	<u>2024 年 1 月 1 日</u>
上 2 桁年	TwoDigitYearOf	<u>2024 年 1 月 1 日</u>
季節	SeasonOf	<u>2024 年冬</u>
四半期	QuarterOf	<u>2024 年第 1 四半期</u>
月	MonthOf	<u>2024 年 1 月</u>
週	WeekOf	<u>2024 年 1 月第 1 週</u>
日	DayOf	<u>2024 年 1 月 1 日</u>
午前・午後	AMPMOf	<u>2024 年 1 月 1 日</u> <u>午前 10 時 11 分 12 秒</u>
時	HourOf	<u>2024 年 1 月 1 日</u> <u>午前 10 時 11 分 12 秒</u>
分	MinuteOf	<u>2024 年 1 月 1 日</u> <u>午前 10 時 11 分 12 秒</u>
秒	SecondOf	<u>2024 年 1 月 1 日</u> <u>午前 10 時 11 分 12 秒</u>

によって、定時間情報表現が得られる。

時間粒度、演算子に関するアノテーションの例は次のとおりである。

例 1 基準日時：2024-08-05, 不定時間情報表現：9 月 3 日, 定時間情報表現：2024-09-03
基準日時を年の粒度までを用いることで欠落した年を補完することができるので、時間粒度タグ YearOf 及び演算 This をアノテーションする。

例 2 基準日時：2024-08-05, 不定時間情報表現：3 日, 定時間情報表現：2024-09-03
基準日時を月の粒度まで取り出し、1 か月を加算した値である「2024-09」を用いて欠落した年、月を補完することができるので、時間粒度タグ MonthOf 及び演算子 After をアノテーションする。

4. タグ、絶対値の推定

推定は、時間表現タグと関係抽出の 2 種類に対して行わ

*2 基準日時へのアンカーは TimeML においても anchorTimeID としてアノテーションされている。BCCWJ-Timebank では少数の TIMEX3 タグの comment 属性にアノテーションされている。

*3 ISO8601-1:2019 における「century」、JIS X 0301:2019 における「百年台」の定義と一致する。

れ、その結果を用いて絶対値の推定が行われる。

本研究では、時間表現タグを固有表現抽出タスク、時間関係の抽出を関係抽出タスクと見なして学習と推定を行う。

関係抽出においては、時間粒度タグと演算子の両方が抽出対象となるが、`This_MonthOf` のように演算子と時間粒度タグを1タグに結合したうえで学習と推定を行う。

絶対値の推定は、時間表現推定器が出力する時間表現のタグ、表層文字列、基準日時から抽出される時間粒度、演算子を引数とし、ルールベースの抽出によって時間区間の推定に必要な始点、終点の各時間粒度の決定を行う。

絶対値の抽出処理の手順を次に示す。

- (1) 表層文字列の Unicode 正規化、漢数字から算用数字への変換を行う。
- (2) 時間表現に区間（例 8～9月）が含まれるかどうか、表層文字列を用いて判定する。区間が含まれる場合は、各時間表現に分割し、分割した各表現に対して以降の処理を行う。
- (3) 日付、時間をルールベースで切り出す。
- (4) 時間表現の情報が不完全である場合は、基準日時、演算子を用いて、不足している時間粒度の情報を補完する。
- (5) 辞書を用いて、主観的な時間表現を絶対値の区間に変換する。
- (6) グレゴリオ暦ではない絶対値をグレゴリオ暦に変換する。
- (7) 始点、終点を出力する。

5. 実験

5.1 データセット

文献 [5] においてアノテーション対象となった記事に対して、本研究で定義したタグ及び関係のアノテーションを行った。対象となっている記事は文献 [5] と同一であり、毎日新聞記事データ集プラス 2010～2019 年版を LDA で分類した結果のうち、地震、洪水のトピック各 100 件、及び、1990 年の全国版記事から日付昇順で選択された 20 件が含まれる。アノテーション作業は、テキストエディタによって時間表現タグと絶対値をアノテーションしたうえで、INCEpTION [8] を用いて関係タグの追加及びアノテーションの修正を行った。INCEpTION におけるアノテーション例を図 1 に示す。

主観的な時間表現は絶対値の時間区間に置換してアノテーションした。置換例を次に示す。

季節 二十四節気の春分、夏至、秋分、冬至を用いて、次のとおり区分を定める。ただし、粒度は日までとする。

春 = [春分, 夏至), 夏 = [夏至, 秋分), 秋 = [秋分, 冬至), 冬 = [冬至, 春分)

初旬、中旬、下旬 対象となる時間粒度を 3 等分して区分を定める。月については日数に関わらず 10 日単位で

区切る。

週初、週半ば、週末 週初 = [月曜日, 水曜日), 週半ば = [水曜日, 土曜日), 週末 = [土曜日, 月曜日) とする。

年始、年末 年始は [XXXX-01-01, XXXX-01-16) とし、年末は [XXXX-12-16, XXXX-01-01) とする。ただし、年末の終点は始点の翌年とする。

時 気象用語を用いて区間を定める。例えば「未明」は [T00, T03) とする。

気象用語において定義されていない表現は独自に定める。例えば、「昼」は [T11, T13) とし、「深夜」は [T21, T03) とする。

5.2 タグ推定、関係抽出の精度評価

構築したデータセットを用いて、タグの推定と関係抽出を行う。モデルの学習とタグの推定には SpERT [9] を用いた。SpERT はスパンに対してタグ推定と関係抽出を行う推定器である。事前学習済みモデルには `nlp-waseda/bigbird-base-japanese`^{*4} を用い、記事の全文を 1 記事単位で入力する。なお、このモデルは Juman++ によって単語分割された単語列によって訓練されているが、ファインチューニングに用いたデータセットは超短単位によってアノテーション基準が策定されており、分割単位を変更せずに入力している。ファインチューニングにおけるパラメータは次のとおり指定した。

- `train_batch_size` = 8
- `epochs` = 100
- `lr` = $2e-5$
- `lr_warmup` = 0.2
- `weight_decay` = 0.01
- `max_grad_norm` = 1.0
- `prop_drop` = 0.3
- `neg_entity_count` = 100
- `neg_relation_count` = 100

時間表現タグ推定と関係抽出の評価指標には、Precision, Recall, F1 を用いる。ただし、関係抽出においては、スパンと時間表現タグの両方を正しく推定し、かつ関係タグを正しく推定できた場合にのみ正解とする。

記事単位で 5-fold cross validation を行った結果の平均値、及び全体の micro 平均、macro 平均を表 3、表 4 に示す。

時間表現タグの推定については、全体的に Recall が高い傾向となった。タグごとの結果では、T タグの F 値が最も高かった。新聞記事の場合、発行日の直近の出来事について述べる内容が多く、月、日、時の時間表現が頻出することが影響したと考えられる。TS タグについては、絶対表現の中では T タグの次に事例数が多いが、Precision の低下が

^{*4} <https://huggingface.co/nlp-waseda/bigbird-base-japanese>

2	太平洋側を中心に降り続いた大雨の影響で、岩手県山田町織笠の高台に造成された仮設住宅の基部が崩れはじめたため、同町は T MonthOf 2012-05-04T01:50 DCT This (Relation) 4 日午前 1 時 50 分、12 世帯 25 人に避難勧告を出した。
3	この約 40 分後、基部はフェンスと共に崩落したが、けが人はなかった。 (Rel...)RTA MinuteOf 2012-05-04T02:30 Link After

図 1 アノテーション例。文書作成日時 (DCT) は 2012-05-05 である。図中の DCT, Link は後処理用の属性であり、基準日時が DCT か文書中のエンティティかを示す。

Fig. 1 Example of annotation. Document Creation Time (DCT) is 2012-05-05. DCT and Link indicates whether the reference is to the DCT or to the entity in the document for post-processing.

表 3 テストセットにおける時間表現タグの推定結果。事例数は全テストセットの合計値である。

Table 3 Test set result of temporal tag prediction. The numbers of examples are the sum over all the test sets.

タグ	Precision	Recall	F1	事例数
T	84.2	96.9	90.1	689
TA	76.0	98.5	85.8	41
TS	68.4	90.2	77.8	144
TSA	70.0	70.0	70.0	8
RT	60.2	94.2	73.4	223
RTA	0.0	0.0	—	4
RTS	100.0	58.3	73.7	12
RTSA	—	—	—	0
micro 平均	75.0	94.6	83.7	
macro 平均	70.8	81.9	75.9	

特徴的である。これについては「3 日夜」などのように TS タグのスパンに含まれる客観的な時間表現が影響したものと考えられる。

時間関係抽出については、時間粒度タグと演算子の結合を 1 タグとすることから、時間表現タグと比べてスパース性が高くなっており、全体的なスコアの低さに影響しているものと考えられる。時間表現タグ同様、時間関係抽出においても全体的に Recall が高い傾向となった。This_MonthOf タグが事例数、F1 値ともに高いが、新聞記事においては日が記事中で言及されやすい傾向にあり、かつ、それらの多くは発行日が基準日となることから、発行日の月までの粒度を取り出すことで定時間情報表現を復元する、というパターンが多いことが影響したと考えられる。

5.3 絶対値推定の精度評価

絶対値推定機のルールセットの構築に当たっては、毎日新聞記事データ集プラス 1990～2019 年版における時間表現アノテーション済みの 12,900 文^{*5}と BCCWJ-Timebank^{*6}において、次の正規化を適用してテキスト中に出現する時間表現の集計を行い、それぞれ 5 件以上事例が存在するものを抽出し、その結果を基に人手で構築を行った。

^{*5} 5.1 で構築したデータセットとの重複はない。

^{*6} type が SET であるものは除外し、DURATION については相対表現と解釈できるものだけを用いた。

(1) 数の置換：数は桁数を維持したまま記号 X に置換する。

例 2024 年→XXXX 年, 12 月→XX 月

(2) 元号の置換：元号は記号<ERA>に置換する。

例 令和 6 年→<ERA>X 年

グレゴリオ暦・和暦間の変換は HuTime Web API (Calendar Calculation) [10] を用いて実装した。

評価は、5.1 において構築したデータセットを推定器に入力し、出力される時間区間が正しいかどうかによって行う。ただし、実験においては絶対表現のみを評価対象とし、「1 日前」などの相対表現は除外した。また、相対表現と間接的に関係をもつ絶対表現も除外した^{*7}。

絶対値推定器には人手アノテーションした時間表現タグ、時間関係タグを入力し、絶対値のみを評価した。評価指標は、TempEval-3 [11] における Attribute Precision, Attribute Recall, Attribute F1 の定義を用いる。この定義では、スパンと時間表現タグが正しく判定され、かつ、始点と終点の時間粒度が正しく推定された場合のみ正解となる。

結果を表 5 に示す。テストセットに対しては、ほぼ全ての表現が正しく推定される結果となった。これは、新聞記事では時間表現のバリエーションが限られており、構築したルールセットで表現を十分に被覆することができたためと考えられる。

推定に失敗した事例は、ルールセットを構築した際に対象外となった少数の事例が該当したもの、表層文字列だけでは時間の型が確定できなかったものであった。推定に失敗した 3 事例を次に示す。

(1) ルールに含まれない表現

元旦 に向け、■■■^{*8} 理事長らが交流のある宮城県石巻市の仮設住宅などに届けることにしている。

推定結果: [2012-01-01, 2013-01-01)

正解: [2012-01-01, 2012-01-02)

(2) 表層文字列だけでは時間の型が判別できない表現

^{*7} 例えば「相対表現—絶対表現—絶対表現」の関係があるとき、下線部のエンティティは相対表現と間接的に関係をもつので除外対象となる。

^{*8} 原文の個人名を伏せ字としている。

表 4 テストセットにおける時間関係抽出の推定結果。事例数は全テストセットの合計値である。

Table 4 Test set result of temporal relation prediction. Number of examples is the sum of all test sets.

演算子	This				Before				After			
粒度	Precision	Recall	F1	事例数	Precision	Recall	F1	事例数	Precision	Recall	F1	事例数
CenturyOf	0.0	0.0	—	2	—	—	—	0	—	—	—	0
EraOf	—	—	—	0	—	—	—	0	—	—	—	0
EraYearOf	—	—	—	0	—	—	—	0	0.0	0.0	—	0
YearOf	32.3	65.8	43.3	173	19.3	65.0	29.8	56	0.0	0.0	—	8
TwoDigit	0.0	0.0	—	37	0.0	0.0	—	6	—	—	—	0
YearOf												
SeasonOf	0.0	0.0	—	1	0.0	0.0	—	5	—	—	—	0
QuarterOf	—	—	—	0	—	—	—	0	0.0	0.0	—	0
MonthOf	79.8	93.5	86.1	450	0.0	0.0	—	9	0.0	0.0	—	2
WeekOf	0.0	0.0	—	2	0.0	0.0	—	1	0.0	0.0	—	0
DayOf	26.6	51.8	35.2	266	0.0	0.0	—	8	0.0	0.0	—	5
AMPMOf	3.3	5.0	4.0	32	—	—	—	0	—	—	—	0
HourOf	0.0	0.0	—	4	0.0	0.0	—	0	0.0	0.0	—	1
MinuteOf	—	—	—	0	0.0	0.0	—	1	0.0	0.0	—	1
SecondOf	—	—	—	0	0.0	0.0	—	0	—	—	—	0
micro 平均	40.5	66.4	50.3									
macro 平均	11.9	20.5	15.0									

表 5 テストセットにおける絶対値の推定結果。事例数は全テストセットの合計値である。

Table 5 Test set result of temporal value prediction. The numbers of examples are the sum over all the test sets.

Precision	Recall	F1	事例数
99.6	99.6	99.6	754

- フォーラム「『復興知』を未来につなぐ～東日本大震災5年を迎えて～」が10、11の両日、西宮市上ヶ原一番町の関西学院大で開かれた。

推定結果: [2016-01-01, 2016-02-01)

正解: [2016-01-10, 2016-01-12)

- 09年の6・30豪雨では、天ヶ瀬温泉街で濁流の一部が堤防からあふれ、旅館など計5棟が床下浸水した。

推定結果: [2009-01-01, 2010-01-01)

正解: [2009-06-30, 2009-07-01)

いずれの例においても、推定器は基準日時を時間粒度で切り出した値を最終的な結果として出力した。

(2)については、それぞれThis_MonthOf, After_YearOfが与えられており、周辺文脈からも日付表現であることが類推可能であるが、スパン内の値の書式や単位を用いて日時の切り出しルールが定義されており、スパン内にはこれらにマッチする文字列がなかったことから正しく推定ができなかったものと考えられる。

5.4 抄録を対象とした時間表現の抽出

他ドメインへの適応の評価として、研究報告の抄録を対象とした時間表現の抽出を行った。

対象は国立歴史民俗博物館研究報告第229集～第248集に含まれる日本語の抄録とし、時間表現タグと絶対値をアノテーションして評価に用いた。

タグ推定、関係抽出に用いる推定器は、5.1で構築したデータセットの全体を用い、同一のパラメータで学習した。また、絶対値推定器には、SpERTが出力したスパンとタグを入力して最終的な絶対値を得た。

評価指標には5.3と同様とする。ただし、このアノテーションデータには関係ラベルが付与されていないことから、相対表現との関係性に基づく除外は行わなかった。

時間表現タグの推定結果を表6に、絶対値の推定結果を表7に示す。全体的に新聞記事のテストセットよりF値が低い結果となった。また、新聞記事とは異なり、このデータセットにおいてはTSタグのF値が最も高い結果となった。これは、TSタグにおいては「～世紀後半」のように、世紀や年と主観的な時間表現とが結合した類似表現が多く出現していることが影響したと考えられる。絶対値の推定においては、時間粒度や演算子のタグの推定誤りによって異なる絶対値に推定された事例が見られた。

6. おわりに

本研究では、主観的な時間表現に注目し、客観的な時間表現と統一的な枠組みで処理を実現するためのアノテー

表 6 抄録における時間表現タグの推定結果

Table 6 Result of temporal tag prediction for abstract.

タグ	Precision	Recall	F1	事例数
T	63.9	69.4	66.5	252
TA	80.0	25.0	38.1	16
TS	84.3	68.3	75.4	63
TSA	0.0	0.0	0.0	2
RT	21.4	66.7	32.4	9
RTA	—	—	—	0
RTS	0.0	0.0	0.0	4
RTSA	—	—	—	0
micro 平均	63.7	65.9	64.8	
macro 平均	41.6	38.2	39.8	

表 7 抄録における絶対値の推定結果

Table 7 Result of temporal value prediction for abstract.

Precision	Recall	F1	事例数
61.0	62.9	62.0	333

ション基準の提案とそれに基づくアノテーションデータの構築及び精度評価を行った。

本研究で策定したアノテーション基準は連続する区間を前提としたものとなっている。したがって、例えば「05～06年秋冬」のように複数の不連続な時間区間の集合は対象外となっている。また、「東日本大震災から13年後」のように、イベントに付随する日付表現を基準日時とする時間表現についても、イベントそのものは時間表現ではないことから対象外となっている。今後、多様なテキストを対象にアノテーションを行うことで、対象とすべき時間表現について検討を行う必要がある。

絶対値の推定においては、入力されるスパンとラベルを利用して高い精度で絶対値を推定できているが、絶対値の推定に基準日時の情報が必要な場合、最終的な絶対値の推定に大きく影響を与えることとなる。よって、スパンとタグの推定精度がより重要となる。

また、絶対値推定のためのルールセットは日本語の表現に特化したものであり、日本語に依存した処理が実装されている。他ドメインへの適応、他言語対応のためにはルールセット及び処理手順の改善が必要となる。また、多様なテキストから時間表現を抽出するためにはルールセットの増加やルールの複雑化が起こる可能性があることから、ルールセットのメンテナンス性の向上も今後の課題となる。

今後は、時間表現タグを眼瞼に推定できるよう、多様なテキストを対象としてアノテーションを行うとともに、アノテーションデータからのルールセットの自動獲得、機械学習を用いた手法などに取り組んでいきたい。

参考文献

[1] Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G. and Radev, D. R.:

TimeML: Robust specification of event and temporal expressions in text., *New directions in question answering*, Vol. 3, pp. 28–34 (2003).

[2] 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, 自然言語処理, Vol. 20, No. 2, pp. 201–221 (2013).

[3] 矢田峻太郎, 田中リベカ, Cheng, F., 荒牧英治, 黒橋慎夫: 汎用的な臨床医学テキストアノテーション仕様およびガイドラインの策定: 重篤肺疾患ドメインに着目して, 自然言語処理, Vol. 29, No. 4, pp. 1165–1197 (オンライン), DOI: 10.5715/jnlp.29.1165 (2022).

[4] Bethard, S. and Parker, J.: A Semantically Compositional Annotation Scheme for Time Normalization, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J. and Piperidis, S., eds.), Portorož, Slovenia, European Language Resources Association (ELRA), pp. 3779–3786 (2016).

[5] Sung, J., Mori, S., Kameko, H., Kubo, A. and Sekino, T.: Inference of Absolute Time Value from Temporal Expressions, *2021 IEEE International Conference on Big Data (Big Data)*, pp. 2273–2280 (online), DOI: 10.1109/BigData52589.2021.9671863 (2021).

[6] Maekawa, K., Yamazaki, M., Maruyama, T., Yamaguchi, M., Ogura, H., Kashino, W., Ogiso, T., Koiso, H. and Den, Y.: Design, Compilation, and Preliminary Analyses of Balanced Corpus of Contemporary Written Japanese, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. and Tapias, D., eds.), Valletta, Malta, European Language Resources Association (ELRA) (2010).

[7] Ding, W., Chen, J., Li, J. and Qu, Y.: Automatic rule generation for time expression normalization, *Findings of the Association for Computational Linguistics: EMNLP 2021* (Moens, M.-F., Huang, X., Specia, L. and Yih, S. W.-t., eds.), Punta Cana, Dominican Republic, Association for Computational Linguistics, pp. 3135–3144 (online), DOI: 10.18653/v1/2021.findings-emnlp.269 (2021).

[8] Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E. and Gurevych, I.: The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation, *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, pp. 5–9 (2018). Event Title: The 27th International Conference on Computational Linguistics (COLING 2018).

[9] Eberts, M. and Ulges, A.: Span-based joint entity and relation extraction with transformer pre-training, *ECAI 2020*, IOS Press, pp. 2006–2013 (2020).

[10] 関野樹: 暦に関する Web API – 暦法の変換と期間の計算, じんもんこん 2017 論文集, pp. 23–28 (2017).

[11] 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)* (Manandhar, S. and Yuret, D., eds.), Atlanta, Georgia, USA, Association for Computational Linguistics, pp. 1–9 (2013).