Comprehensive Annotation of Various Types of Temporal Information on the Time Axis

Tomohiro Sakaguchi, Daisuke Kawahara and Sadao Kurohashi
Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
sakaguchi@nlp.ist.i.kyoto-u.ac.jp, {dk, kuro}@i.kyoto-u.ac.jp

Abstract

In order to make the temporal interpretation of text, there have been many studies linking event and temporal information, such as temporal ordering of events and timeline generation. To train and evaluate models in these studies, many corpora that associate event information with time information have been developed. In this paper, we propose an annotation scheme that anchors expressions in text to the time axis comprehensively, extending the previous studies in the following two points. One of the points is to annotate not only expressions with strong temporality but also expressions with weak temporality, such as states and habits. The other point is that various types of temporal information, such as frequency and duration, can be anchored to the time axis. Using this annotation scheme, we annotated a subset of Kyoto University Text Corpus. Since the corpus has already been annotated predicate-argument structures and coreference relations, it can be utilized for integrated information analysis of events, entities and time.

Keywords: Time annotation, Time anchoring, Time axis

1. Introduction

Everyday many texts are generated on the Web, and a huge amount of texts have been accumulated so far. To extract knowledge about a certain topic from this large amount of texts, we need an information analysis technology to integrate, summarize and compare related texts. In order to analyze texts written at different times or texts referring to different times, it is necessary to interpret the temporal information implied in the texts. There have been many studies and tasks to understand the relationship between event information and time information in text. For example, temporal ordering of events that estimates the temporal relations of event-event and event-time was studied in TempEval 1, 2, 3 (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013), and the timeline generation task that links event and time in multiple documents was studied in SemEval 15 (Minard et al., 2015).

In order to train models and evaluate results in these tasks, corpora in which event information is correlated with time information in text have been developed (Pustejovsky et al., 2003; Cassidy et al., 2014; Reimers et al., 2016). In these studies, expressions which have clear temporality were annotated, but in order to know how people understand texts from the perspective of time, it is essential to know how the expressions with weak temporality are interpreted. To understand temporal information in text exhaustively, we propose an annotation scheme that anchors various expressions to the time axis, reflecting personal interpretation of text and common sense. Using this scheme, we annotate Kyoto University Text Corpus (Kawahara et al., 2002), which is a Japanese newspaper corpus annotated with predicate-argument structures and coreference relations.

The points of our annotation scheme are two-fold. One of the points is to annotate various expressions that can have temporality. We annotate not only expressions with strong temporality but also expressions with weak temporality. Many previous studies annotate "events" that express situations that happen or occur, which are defined in the guideline of TimeML (Sauri et al., 2006). Therefore, expressions as in the following example are not annotated.

(1) *Businesses are **emerging** on the Internet so quickly that no one, including government regulators, can keep track of them.*

However, the temporal information of expressions other than "event" also can be a clue to understand text. In the case of the above example, the temporal information of *"emerging,"* i.e., several years ago to the present, should be annotated to clarify the temporal common sense implied in the text. Therefore, we annotate all the expressions that can have temporality, that is, all the predicates and the eventive nouns in text. Annotators judge whether the expressions have temporality, and annotate the corresponding time tags. The other point of our annotation scheme is that various types of time information such as frequency and duration can be anchored to the time axis. Reimers et al. (2016) proposed an annotation scheme that represents an event period using its starting and ending points. However, it cannot represent "non-continuous period" or "a period in a long duration" as in the following examples.

(2) *He **plays** baseball every Sunday.*

(3) *I will **take** a business trip for three days next week.*

(4) *He often used to **have** a tea with us.*

In this paper, we introduce new time tags that can more accurately anchor various types of time information to the time axis.

By annotating various types of temporal information with the expressive time tags, personal interpretation of text and common sense appear as tag disagreements. In this research we consider that such disagreements are also important in understanding how time information is interpreted, and thus we do not eventually integrate time tags annotated by several annotators into one. Instead, we introduce an annotation method that keeps differences in interpretation and only corrects obvious annotation errors.

Using the annotation scheme, we annotated 113 documents with 4,534 expressions in Kyoto University Text Corpus. 80% of the expressions are judged to have temporality, and approximately 30% of them are annotated with the notation newly proposed in this paper. Since the corpus has already been annotated with predicate-argument structures and coreference relations, our annotation makes it possible to utilize for integrated information analysis of events, entities and time. The annotated corpus is publicly available.[1]

2. Related Work

There are many corpora which associate event information with time information, and they can be roughly divided into two approaches. One approach is annotating temporal relations between events. Pustejovsky et al. (2003) annotated events and times based on the TimeML guideline, and relations between event-event, event-time, and event-time. Originally, the annotation was sparse because there were only the relations which are judged to be important by annotators, but TempEval competitions (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013) annotated all the relations in same sentence to improve the coverage.

Kolomiyets et al. (2012) annotated temporal order relations with the nearest event expressions in a corpus of children's stories. Cassidy et al. (2014) annotated all temporal relations in the same sentence and neighbouring sentences.

The other approach is anchoring events to the time axis. Huang et al. (2016) annotated one of five temporal status categories with events in newspaper articles on civil unrest: *Past, On-going, Future Planned, Future Alert, Future Possible*.

Reimers et al. (2016) anchored with finer granularity. Their smallest granularity is day. They divided events into two types: single day event and multiple day event. The former is annotated with the date on which the event occurred, and the latter is annotated with the start and end dates of the event. For example, ***sent*** in the following sentence, an event which ends in one day, is annotated with *1980-05-26*, and ***spent***, an event spanning multiple days, is annotated with *beginPoint=1980-05-26 endPoint=1980-06-01*.

(5) He was ***sent*** into space on May 26, 1980. He ***spent*** six days aboard the Salyut 6 spacecraft.

In the case that the exact event date is not mentioned, notations *before* and *after* are used. In the following sentence, ***appointed*** is annotated with *after 1996-01-01 before 1996-12-31*, and ***part*** is annotated with *beginPoint=after 1984-10-01 before 1984-10-31 and endPoint=after 1984-10-01 before 1984-10-31*.

(6) In 1996 he was ***appointed*** military attache at the Hungarian embassy in Washington. [...] McBride was ***part*** of a seven-member crew aboard the Orbiter Challenger in October 1984

In their annotations, about 60% of all the events end in a day, and about 40% are events that span multiple days. 56% of the former have precise dates, and of the latter, 20% have precise start dates and 16% have precise end dates.

In this research, we extend the anchoring to the time axis approach, and propose annotation scheme that can deal with various time information in text.

3. Annotation Scheme

We annotate expressions which consist of all predicates and eventive nouns in text (hereinafter referred to as "target expressions"). We first apply morphological analysis to text and extract base phrases of predicates and eventive nouns. Annotators first judge whether the expressions have temporality. Expressions that have temporality are annotated with time tags which represent the corresponding time value in consideration of the document creation time (DCT) and the context. When an expression is judged to have no temporality, it is annotated with the time tag *not applicable (t:n/a)*.

A time tag that has temporality is expressed as a Time Base Unit (TBU) or a combination thereof (Table 1). TBU represents a specific time point, and TBUs are divided into three types. There are four ways of combining these, which enable to represent various types of time information. Reimers et al. (2016) used only TBU 1 and combinations a and c in Table 1. As in the previous studies, the finest granularity of time tags is day.

Although we annotate the Japanese corpus, we use English examples to explain our annotation scheme below.

3.1. Judgement of Temporality

In order to judge the temporality of a target expression, annotators consider whether it implies a change in the behavior or state between the past and future. In the case that the expression focuses on a change, it has temporality. In the following examples, ***go*** in example (7) and ***thriving*** in example (8) have temporality, and ***eat*** in example (9) has no temporality. Note that ***thriving*** in example (8) is an expression which is not subject to annotate in the previous studies.

(7) *He will **go** to Kyoto tomorrow.*

(8) *Language processing research is **thriving**.*

(9) *Rabbits **eat** grass.*

3.2. Time Base Unit (TBU)

3.2.1. Date Tag

The temporal information of a date expression is represented by annotating the time value in *t* tag. The time value notation in Japanese TimeBank Corpus (Asahara et al., 2014) is used, such as *t:YYYY* and *t:YYYY-MM-DD*. To reduce the annotation cost, the date tag of the document creation time can be written as *t:DCT*.

For example, ***arrived*** in the following sentence is annotated with *t:2017-04-28*.

(10) [DCT: 2017-04-29] *The president **arrived** in New York yesterday.*

Unlike the previous studies, our annotation scheme allows time tags with larger granularity than day. For example, ***hot*** in the following example is annotated with *t:2016-08*.

[1] http://nlp.ist.i.kyoto-u.ac.jp/index.php?KUTBC

Temporality	Time Base Unit (TBU)	1. Date tag (Day, Month, Year)	e.g., *t:1995-01-05*
		2. Vague time tag (Past, Present, Future)	e.g., *t:PRESENT*
		3. Relative tag (Time Coreference, Utterance Day)	e.g., *t:election*
	Combinations of TBUs	a. Interval between TBUs (TBU∼TBU)	e.g., *t:1995-01-05∼1995-01-07*
		b. Specific span in a TBU (span)	e.g., *t:1995-01,span:P1W*
		c. Unspecific span in a TBU (partial span)	e.g., *t:1995-01,span:part*
		d. Repetition of TBUs (freq)	e.g., *t:1995-01,freq:2/P1W*
No Temporality			e.g., *t:n/a*

Table 1: List of time tags

(11) [DCT: 2017-04-29] *It was **hot** last August.*

This tag does not necessarily mean exactly from August 1st to 31st. As the expression "August" is different from "August 1st to 31st," the period of the corresponding time tag is somewhat vague. The granularity of the time tags in this paper implies such vagueness.

3.2.2. Vague Time Tag

There are many expressions that represent vague time in text. In the following sentence, it is not clear when and how long ***live*** represents in the past.

(12) *I used to **live** in Hiroshima.*

Reimers et al. (2016) interpreted this expression as "a period from one day to another day until today" and annotated it with *beginPoint=before DCT endPoint=before DCT*. In our annotation scheme, some special tags are introduced. The vague past, present and future are represented as *t:PAST*, *t:PRESENT* and *t:FUTURE*, respectively. *t:PRESENT* includes not only today but also a little past and future. In the following sentence, ***bring*** is annotated with *t:PRESENT* since it represents not only today but also a little before and after today.

(13) *You can **bring** liquids on domestic flights.*

To represent the past and future, *t:PAST-M*, *t:PAST-Y*, *t:FUTURE-M* and *t:FUTURE-Y* tags are also available according to the temporal distance. *t:PAST-M* represents a few months ago and *t:PAST-Y* represents a few years ago. For more than a few years ago, or when the granularity is unknown, *t:PAST* is used. It is the same for future.
There are other vague time expressions. In the case of expressions that represent numerical ambiguity, such as "around 1980" or "about 3 years", *ap* (approximately) is attached to the ambiguous numerical value of the time tag. In the following sentence, ***built*** is annotated with *t:1980ap*.

(14) *The hotel was **built** around 1980.*

3.2.3. Relative Tag

In texts with few temporal expressions, such as novels, it is difficult to anchor events to the time axis. In such a case, the TimeBank Corpus' annotation scheme, i.e., annotating the temporal relation between events, provides richer information. Therefore, in the case where the specific date is unknown but the temporal relation with another phrase in the same sentence is known, that phrase is used as a time value (*Time Coreference*). In the following sentence, though the date on which the demonstration took place is unknown, it can be understood that it is the day after the election. In this case, ***held*** is annotated with *t:election+P1D*, using the notation of the duration expressions in TimeBank Corpus (see subsection 3.3.2. for details).

(15) *The day after the election, a large demonstration was **held**.*

If there are two or more phrases that can be referred to, priority is given as follows and one with the highest priority is selected: 1. phrase with absolute time value tag, 2. phrase with the closest distance.
In conversational sentences and interviews, the date of the speech is often unknown. If the date of the utterance cannot be guessed from the context, the date can be described as *t:UD* (*Utterance Day*). In the following sentence, ***work*** is annotated with *t:UD*. Note that ***said***, an expression outside the utterance, is annotated with the absolute time value.

(16) *"I have no choice but to **work** hard from now," said the director.*

3.3. Combinations of TBUs

3.3.1. Interval between TBUs

The interval between TBUs is represented by connecting the starting TBU and the ending TBU with ∼. This notation corresponds to the *beginPoint* and *endPoint* tags in Reimers et al. (2016). If either of the starting or ending TBU cannot be guessed, it is omitted. The time tag of ***busy*** in the following sentence is *t:∼2017-04-28*.

(17) [DCT: 2017-04-29] *I was **busy** up until yesterday.*

3.3.2. Span in a TBU

A part of the period in a long TBU, e.g., a part of the period in August, is represented by combining the *t* tag representing the large period and the *span* tag representing the small period. When the length of the small period is guessed, the *span* tag is represented using the notation of the duration expressions defined in the Japanese TimeBank Corpus. For example, three years is represented as *span:P3Y*, three weeks is represented as *span:P3W* and three days is represented as *span:P3D*. In the following sentence, ***am going*** is annotated with *t:2017-05,span:P1W*.

(18) [DCT: 2017-04-29] *I **am going** to London for a week next month.*

If the length of the small period cannot be guessed, it is represented as *span:part*. The *span:part* tag is equivalent to the *before* and *after* tags in Reimers et al. (2016).

	Annotator1	Annotator2	Annotator3	Average
Date tags	1195 (26.4%)	1145 (25.3%)	938 (20.7%)	1093 (24.1%)
Year	35 (0.8%)	47 (1.0%)	16 (0.4%)	33 (0.7%)
Month	9 (0.2%)	14 (0.3%)	3 (0.1%)	9 (0.2%)
Day	1151 (25.4%)	1084 (23.9%)	919 (20.3%)	1051 (23.2%)
Vague time tags ∗	617 (13.6%)	375 (8.3%)	249 (5.5%)	414 (9.1%)
t:PRESENT	520 (11.5%)	257 (5.7%)	195 (4.3%)	324 (7.2%)
t:PAST	40 (0.9%)	25 (0.6%)	17 (0.4%)	27 (0.6%)
t:FUTURE	57 (1.3%)	89 (2.0%)	31 (0.7%)	59 (1.3%)
t:ap	0 (0.0%)	4 (0.1%)	6 (0.1%)	13 (0.3%)
Relative tags ∗	215 (4.7%)	135 (3.0%)	313 (6.9%)	410 (8.0%)
Tags including Time Coreference	138 (3.0%)	58 (1.3%)	207 (4.6%)	134 (3.0%)
Tags including *t:UD*	77 (1.7%)	77 (1.7%)	106 (2.3%)	87 (1.9%)
Interval between TBU (∼)	387 (8.5%)	562 (12.4%)	842 (18.6%)	597 (13.2%)
Specific span in a TBU (span) ∗	540 (11.9%)	447 (9.9%)	550 (12.1%)	512 (11.3%)
Date + 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%)
Vague + span	12 (0.3%)	21 (0.5%)	20 (0.4%)	18 (0.4%)
Unspecific span in a TBU (span:part)	455 (10.0%)	561 (12.4%)	478 (10.5%)	498 (11.0%)
Date + 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%)
Vague + span:part	46 (1.0%)	30 (0.7%)	41 (0.9%)	39 (0.9%)
Repetition of TBU (freq)	46 (1.0%)	52 (1.2%)	47 (1.0%)	48 (1.1%)
No temporality (*t:n/a*)	1071 (23.6%)	1077 (23.8%)	1060 (23.4%)	1069 (23.6%)
Tags marked in the second step	8 (0.2%)	180 (4.0%)	57 (1.3%)	82 (1.8%)
Newly proposed tags	1372 (30.3%)	957 (21.1%)	1112 (24.5%)	1147 (25.3%)
All	4534	4534	4534	4534

Table 2: Distribution of annotated time tags. Time tags with ∗ are newly proposed in this paper.

3.3.3. Repetition of TBU

There are many target expressions that are not represented as continuous periods, such as "every Sunday" and "once every three days." Target expressions occurring across multiple days repeatedly are represented with *freq* tag.
The *freq* tag can be used in three ways.

1. When the repetition is expressed as a number of occurrences during a certain period, such as "twice a week" and "once every three days," the *freq* tag is represented as *the number of times / period*. In the following sentence, ***go*** is annotated with *t:2016-07∼DCT,freq:2/P1W*.

   (19) [DCT: 2017-04-29] *I **go** to the pool twice in a week since July 2016.*

2. When the repetition is expressed as a repetition of specific date, such as "every 25th day" and "every Sunday," the date is used as a value of the *freq* tag. The Japanese TimeBank Corpus' notation is extended by allowing to include the symbol @ in each part of *YYYY-MM-DD* in the sense that it can represent any number. In the following sentence, ***is held*** is annotated with *t:PRESENT,freq:@@@@-@@-25*.

   (20) [DCT: 2017-04-29] The Tenjin market ***is held*** on the 25th of every month.

3. When the repetition or the frequency cannot be guessed from the context, one of the following four abstract tags is used: *usually, often, sometimes* and *rarely*. In the following sentence, ***go*** is annotated with *t:PRESENT,freq:sometimes*.

   (21) [DCT: 2017-04-29] I sometimes ***go*** to Starbucks.

	Strict	Relax
The first step	0.417	0.719
The final result	0.554	0.802
The first step (Excluding *t:n/a*)	0.380	0.803
The final result (Excluding *t:n/a*)	0.526	0.867
[Reimers+ 16]	0.617	0.912

Table 3: Inter-annotator agreement computed by Krippendorff's α.

4. Annotation Study

4.1. Annotation Method

Using our annotation scheme, we annotated a subset of documents in Kyoto University Text Corpus. The subset consists of 4,534 target expressions in 113 documents. Kyoto University Text Corpus is a Japanese newspaper corpus that was manually annotated with various linguistic information, such as predicate-argument structures and coreference relations.
The time tags were annotated by three annotators. Since we annotate expressions whose interpretation varies depending on the individual's common sense, we do not eventually combine the annotators' tags into one. We introduce a two-step annotation method that keeps the interpretation

Agreed between annotators			Disagreed between annotators		
Pair of abstracted time tags		Frequency	Pair of abstracted time tags		Frequency
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
All		2045	All		2275

Table 4: Frequency of agreed/disagreed time tags in the first step in the strict metric

Agreed between annotators			Disagreed between annotators		
Pair of abstracted time tags		Frequency	Pair of abstracted time tags		Frequency
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
All		3533	All		787

Table 5: Frequency of agreed/disagreed time tags in the first step in the relaxed metric

of other annotators and modifies only obvious annotation errors. The document set is divided into three parts. Each annotator annotates two of them in the first step, and the remaining one is annotated in the second step. In the first step, each annotator independently annotates, and in the second step they annotate tags by confirming the others' tags in the first step. If an obvious error is found in the already annotated tags, it is just marked. The marked tags are 2% of the total and are treated as missing values in the analysis in section 5.

4.2. Distribution of the Annotated Time Tags

The distribution of the annotated time tags is shown in Table 2. Approximately 80% of the target expressions have temporality, and 55% of them are TBU and the others are combinations of TBUs. The date tags account for approximately 25%, while the vague time tags and the relative tags account for approximately 10%. Since the domain of annotation is newspaper, the majority of target expressions are directly anchored to the time axis. The *freq* tag, representing repetition, is hardly used, i.e., 1% of the whole. The time tags that are newly proposed in this paper account for 25% of the whole.

4.3. Inter-Annotator Agreement

We compute the inter-annotator agreement using Krippendorff's α (Krippendorff, 2004; Hayes and Krippendorff, 2007). Following Reimers et al. (2016), two metrics are utilized. One is a strict metric that measures whether the time tags completely match. The other is a relaxed metric that permits partial matching. If the time tags are overlapped even for one day, they are regarded as matched, and if they do not overlap at all, they are regarded as mismatched. Table 3 shows the agreement at each step. "Excluding *t:n/a*" means an agreement computed excluding the expressions in which one or more annotators annotated with *t:n/a*.

Comparing the first step and the final result of the annotation process, the latter agreement increased significantly. This is because while the documents are annotated independently in the first step, annotators can check others' tags in the second step. When the target expressions annotated with *t:n/a* are excluded, the relaxed agreement increased significantly. It shows that the difficulty of temporality judgement is a cause of lowering the agreement in relaxed metric. Compared with previous studies, the agreement in the strict metric is particularly low. Due to the increase of the variation of the time tags, annotators' interpretations can be reflected a lot, and it became difficult to agree completely.

5. Disagreement Analysis

In order to analyze the annotated time tags without being limited to specific values, we abstract them from the aspect of granularity. For example, for the *t* tag, *t:1994-12-31* is abstracted as *DAY*, *t:∼1994-12-31* is abstracted as *∼DAY* and *t:1994* is abstracted as *YEAR*. For the *span* tag and the *freq* tag, their values are omitted. For example, *t:∼1994-12-31,span:P1D* and *t:∼1994-12-31,span:part* are both abstracted as *∼DAY,span*.

In this section, we analyze the results of the first stage, where annotators independently annotated. Tables 4 and 5 show disagreements in the strict and relaxed metrics respec-

tively. Table 4 indicates that in the strict metric, about 70% of agreed tags are *DAY* and *n/a*, and most of the disagreements are the judgement of temporality and the interpretation of date and period such as *DAY* and ∼*DAY*. Table 5 indicates that most of the disagreements in the relaxed metric are the judgement of temporality. It indicates that most of the tags that were disagreed due to the interpretation between date and period in the strict metric overlap the spans, and they are consistent in the relaxed metric.
In the following subsection, we analyze the disagreement of temporality judgment and the disagreement of interpretation of the date and the period with actual examples.

5.1. Judgement of Temporality

In the relaxed metric, the biggest cause of disagreements is that the judgement of temporality varies depending on annotators. When one annotator tags *n/a*, the other annotates *n/a* (76.6%), *DAY* (5.3%), *PRESENT* (5.0%), ∼*DAY,span* (1.7%) in order of frequency. This means that 75% of *n/a* tags agree, and if it is not the case, one annotates the *DAY* or *PRESENT* tag at a rate of 40%. Many of these expressions represent states, positions and organizations, and the judgment is divided according to whether it is interpreted as permanent or as a temporal period.
In the following sentence, one annotated *t:PRESENT* and the other annotated *t:n/a*.

(22) 大統領官邸の**ある**中心部
The city center where the presidential official residence ***exists***

The annotator who recognized temporality interpreted that there is a possibility that the place of the presidential office may change in the future, while the other interpreted it as semi-permanent.

5.2. Interpretation of Date and Period

As Reimers et al. (2016) pointed out, it is difficult to judge whether an event ends in one day or is held for several days from a text. It is also not easy to clarify the beginning and ending date of an event. Such vagueness appears as disagreements among *DAY*, ∼*DAY*, ∼*DAY,span*, *DAY*∼, *DAY*∼*,span* and *PRESENT* in this annotation scheme. Among them, the disagreement between *DAY* and ∼*DAY,span* often occurs. In many cases, *DAY* is DCT, which means that it is difficult to interpret whether it occurred at the written date or before that.
In the following sentence, it is difficult to judge the duration of the event ***resists*** from the text. One annotated *t:DCT* and the other annotated *t:*∼*DCT,span:part*. While the former interpreted that the event occurred in a day, the latter interpreted as a longer period.

(23) しかしドゥ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。
But the Dudaev regime strongly ***resists***, and the death toll will reach hundreds.

One of the difficulties is due to the domain being newspaper. In the following sentence, one annotated *t:DCT* and the other annotated *t:*∼*DCT,span:P1D*. While the former interpreted that it happened on the date when the article was written from the promptness of newspaper, the latter interpreted that it was not necessarily so.

(24) 外相は、「非民営化・再国営化」の基本方針を**打ち出した**。
The Foreign Minister ***has laid out*** the basic policy of "privatization and re-nationalization."

Thus, the major cause of the disagreements among the annotators is that there are multiple interpretations depending on the context and common sense, closely related to the writing style and theme of newspaper.

6. Conclusion

In this paper, we described a new annotation scheme for comprehensively annotating temporal information in texts reflecting personal interpretation and common sense. Using this scheme, we annotated a subset of a Japanese newspaper corpus, and the new tags account for approximately 25% of the all tags.
Though we annotated newspaper articles in this research, it seems that the writing style and its character are one of causes of the annotation disagreement. In the future, we would like to try annotating on a corpus other than newspaper such as Web texts.

Acknowledgments

This work was partially supported by JST CREST Grant Number JPMJCR1301 including AIP challenge program, Japan. We also thank Manami Ishikawa, Marika Horiuchi and Natsuki Nikaido for their careful annotation.

7. References

Asahara, M., Kato, S., Konishi, H., Imada, M., and Maekawa, K. (2014). BCCWJ-Timebank: Temporal and Event Information Annotation on Japanese Text. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 19, Number 3, September 2014*.

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)*, pages 501–506, Baltimore, Maryland, June. Association for Computational Linguistics.

Hayes, A. F. and Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1):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*, pages 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 Third International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands - Spain, May. 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, The 50th annual meeting of the association for computational linguistics, Jeju, Republic of Korea, 8-14 July 2012*, pages 88–97. ACL.

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)*, pages 778–786, Denver, Colorado, June. Association for Computational Linguistics.

Pustejovsky, J., Hanks, P., Saurí, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Beth Sundheim, D. D., Ferro, L., and Lazo, M. (2003). The timebank corpus. In *Proc. Corpus Linguistics 2003*, pages 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)*, pages 2195–2204, Berlin, Germany, August. Association for Computational Linguistics.

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 *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 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 Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 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*, pages 57–62. Association for Computational Linguistics.