BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text

Masayuki Asahara **Sachi Yasuda** **Hikari Konishi**
Mizuho Imada **Kikuo Maekawa**
Center for Corpus Development, National Institute for Japanese Language and Linguistics
10-2, Midori-cho, Tachikawa, Tokyo, Japan, 190-8561
masayu-a@ninjal.ac.jp

Abstract

Temporal information extraction can be split into the following three tasks: temporal expression extraction, time normalisation, and temporal ordering relation resolution. This paper describes a time expression and temporal ordering annotation schema for Japanese, employing the *Balanced Corpus of Contemporary Written Japanese*, or BCCWJ. The annotation is aimed at allowing the development of better Japanese temporal ordering relation resolution tools. The annotation schema is based on an ISO annotation standard – *TimeML*. We extract verbal and adjective event expressions as ⟨EVENT⟩ in a subset of BCCWJ. Then, we annotate temporal ordering relation ⟨TLINK⟩ on the above pairs of event and time expressions by previous work. We identify several issues in the annotation.

1 Introduction

Temporal information processing in natural language texts has received increasing scholarly attention in recent years. Since temporal order of events often has implications for causal relations (*cause and effect*), identifying them is an essential task for deep understanding of language. Several types of resource for English temporal information processing have been developed, such as an annotation specification *TimeML* (Pustejovsky et al., 2003) and annotated corpora *TimeBank* (Pustejovsky et al., 2010) and *Aquaint TimeML Corpus*. The English annotation specification has been extended as an ISO standard of a temporal information mark-up language – ISO TimeML (ISO, 2008), which covers Italian, Spanish, Chinese and other languages. Temporal information-annotated corpora in various languages have been developed and shared by natural language processing researchers. TempEval-2 (Verhagen et al., 2010), a task for the SemEval-2010, and TempEval-3 (UzZaman et al., 2013), a task for the SemEval-2013, have been proposed as shared temporal-relation reasoning tasks. In these shared tasks, datasets for English, Italian, Spanish, Chinese, and Korean are provided.

However, there is no such resource for the Japanese language. In this paper, we present a means of porting ISO-TimeML into the Japanese language and also describe the basic specifications of '*BCCWJ-TimeBank*' which is a realisation of the temporal information annotation of the *Balanced Corpus of Contemporary Written Japanese*, or BCCWJ (Maekawa, 2008).

2 Related Research

This section explains two related research areas. One is ISO-TimeML, which is an ISO standard for temporal information mark-up languages. The other is BCCWJ, on which we annotate temporal information tags.

2.1 ISO-TimeML

The ISO Technical Committee (TC 37) proposes several standards for language resources, under the collective category 'Terminology and other language and content resources'. Four structures of the committee (SC) are established: TC 37/SC 4 [1] is charged with looking at annotation standards for all areas of natural language resources. TC 37/SC 4 includes six working groups (WG) to design language annotation specification mark-up languages such as stand-off mark-up and XML. TC 37/SC 4/WG 2, the semantic annotation WG, discusses semantic annotation standards. The original TimeML developers and TC 37/SC

[1] http://www.tc37sc4.org/

27th Pacific Asia Conference on Language, Information, and Computation pages 206－214

4/WG 2 defined ISO-TimeML as Semantic Annotation Framework(SemAF)-Time (ISO-24617-1:2012) within the context of TC 37/SC 4.

TimeML and ISO-TimeML define four types of entities — ⟨TIMEX3⟩, ⟨EVENT⟩, ⟨MAKEINSTANCE⟩, and ⟨SIGNAL⟩. The ⟨TIMEX3⟩ tag specifies various attributes of time expressions, such as tid, type, quant, freq, mod, and value. The time expressions are categorised into four types: DATE, TIME, DURATION, and SET. The attribute @value includes the normalised values of the time expressions in a machine-readable format. The ⟨EVENT⟩ tag specifies various attributes of event expressions, including the class of the event, tense, grammatical aspect, polarity, and modal information. The ⟨MAKEINSTANCE⟩ tag presents the event instances expressed by ⟨EVENT⟩-tagged expressions. Finally, the ⟨SIGNAL⟩ tag annotates elements to indicate how temporal objects are related amongst themselves.

TimeML and ISO-TimeML also define several types of links. Among these, ⟨TLINK⟩ expresses temporal order among instances of time expressions and/or event expressions.

2.2 BCCWJ

BCCWJ was publicly released in 2011 by NINJAL , Japan. It consists of three sub-corpora: 'Publication', 'Library', and 'Special purpose'. 'Publication' consists of samples extracted randomly from the whole body of books, magazines, and newspapers published during 2001-2005. 'Library' consists of randomly extracted samples in circulation in libraries in the period 1986-2005. Finally, the 'Special purpose' sub-corpus consists of several mini-corpora without any statistical sampling method being used. It includes text from Yahoo! Answers, Yahoo! Blogs, white papers, and school textbooks. The total size of BCCWJ is about 100 million words.

The part of BCCWJ called 'CORE' , manually annotates word boundaries, base phrase boundaries, and morphological information. CORE consists of six registers in 'Publication' and 'Special purpose': books (PB), magazines (PM), and newspapers (PN) from 'Publication', and Yahoo! Answers (OC), Yahoo! Blogs (OY), and white papers (OW) from 'Special purpose'. The size of CORE is about 1.3 million words.

CORE has received linguistic annotations from several research institutes (e.g. for syntactic dependency structures, by NAIST and NINJAL; predicate-argument relations, by NAIST, named entities by TITECH , modality, by Tohoku and Yamanashi Universities; Japanese framenet , by Keio University, and so on). The CORE samples are split into annotation priority sets from A to E to allow the annotations to overlie as much as possible. Table 1 shows the basic statistics and priority sets of BCCWJ CORE. The word unit is based on the 'Short Unit Word', UniDic standard (Den et al., 2008); UniDic is a lexicon for Japanese morphological analysis.

3 Specification for Japanese Temporal Information Annotation

This section presents a specification for Japanese temporal information annotation. The annotation is realised as BCCWJ-TimeBank. The specification is based on TimeML (Pustejovsky et al., 2003) and adapted to the Japanese language. Figure 1 shows an example of the annotation. Below, we overview the specification of TimeML tags: ⟨TIMEX3⟩ for temporal expressions, ⟨EVENT⟩ and ⟨MAKEINSTANCE⟩ for event expressions, and ⟨TLINK⟩ for temporal ordering. We also mention other tags which we exclude from Japanese temporal information annotation.

3.1 ⟨TIMEX3⟩

The target temporal expressions of ⟨TIMEX3⟩ are DATE, TIME, DURATION, and SET by @type. We do not permit any nests of ⟨TIMEX3⟩ . We clip the expressions by character-based since Japanese does not have word delimitation spaces.

The attributes of @tid, @type, @value, @freq, @quant, and @mod have been inherited from the original TimeML.

There is an issue regarding which calendar to use in porting TimeML to Japanese. In Japan, we use not only the Western calendar but also a native Japanese calendar based on the year of the Emperor's reign. We introduce a new attribute @valueFromSurface to address this issue. @valueFromSurface includes a @value-like like string to indicate a machine-readable datetime value, whereas @value includes the normalised version of value, @valueFromSurface includes the non-normalised version of the value, which can be generated on rewrite rules. @valueFromSurface

PN23_00001 Sample in BCCWJ CORE

```
<TIMEX3 @value="2002-04-11" @definite="true" @tid="t0"
functionInDocument="CREATION_TIME" type="DATE"/>

<sentence>　地方自治体が<EVENT @class="NULL" @eid="e25">運営する</EVENT>公営地下鉄二十六路線のうち
<TIMEX3 @value="FY2000" @definite="FALSE" @valueFromSurface="FY2000" @tid="t4" @type="DATE">
二〇〇〇年度</TIMEX3>決算で経常損益が黒字なのは、
札幌市南北線など四路線に<EVENT @class="I_ACTION" @eid="e26">とどまった</EVENT>ことが、公営交通事業協会
が
<TIMEX3 @value="2002-04-10" @definite="true" @valueFromSurface="XXXX-XX-10" @tid="t5"
type="DATE">十日</TIMEX3><EVENT @class="I_ACTION" @eid="e27">まとめた</EVENT>報告書で
<EVENT @class="I_STATE" @eid="e28">分かった</EVENT>。</sentence>

<MAKEINSTANCE @eventID="e26" @eiid="ei26"/>
<MAKEINSTANCE @eventID="e27" @eiid="ei27"/>
<MAKEINSTANCE @eventID="e28" @eiid="ei28"/>

<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="during" @task="DCT"
@timeID="t0" relatedToEventInstance="ei26"/>
<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="after" @task="DCT"
@timeID="t0" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="after" @task="DCT"
@timeID="t0" @relatedToEventInstance="ei28"/>

<TLINK @relTypeA="vague" @relTypeB="equal" @relTypeC="during" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei26"/>
<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="before" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="before" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei28"/>
<TLINK @relTypeA="after" @relTypeB="before" @relTypeC="during" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei26"/>
<TLINK @relTypeA="contains" @relTypeB="after" @relTypeC="finishes" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="contains" @relTypeB="equal" @relTypeC="before" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei28"/>

<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="contains" @task="E2E"
eventInstanceID="ei26" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="before" @relTypeB="before" @relTypeC="before" @task="E2E"
eventInstanceID="ei27" @relatedToEventInstance="ei28"/>
```

English Translation:
Municipal Transportation Works Association published a report on April 10th. The report shows that only 4 public tube railways (e.g. Sapporo City Nanboku line) from 26 have a surplus.

Figure 1: An Example of Japanese BCCWJ TimeBank annotation

can encode Japanese calendars. For example, '平成 25 年' (*the 25th year of the Heisei era*) is encoded in the `@valueFromSurface` of '`H25`' and normalised as the `@value` of '`2013`' in the ISO-8601-like format.

The difference between `@value` and `@valueFromSurface` shows the use of the normalisation procedure. However, we cannot judge whether the ⟨`TIMEX3`⟩ is fully normalised (fully specified) or under-specified. We introduce another new attribute `@definite` to indicate whether the ⟨`TIMEX3`⟩ is fully-specified '`true`' or under-specified '`false`'.

3.2 ⟨`EVENT`⟩ and ⟨`MAKEINSTANCE`⟩

Next, we need to annotate the event expressions and instances to link the temporal ordering to ⟨`TIMEX3`⟩.

The event expression candidates are automatically extracted from the BCCWJ of morphological information. We define long word units with verbs and adjectives — 4,953 expressions — as the event expression candidates. First, the candidates are judged by two annotators as to whether the target expression is an event expression or not. If the expression boundaries are not valid, a longer expression covering the candidate is redefined by the annotators. Second, the annotators judge whether the target expres-

Table 1: BCCWJ CORE: Registers and its priority set

Register	(Abbr.)	Priority Set	# of Samples	# of Words
White Paper	OW	A to D	62	197,011
Books	PB	A to D	83	204,050
Newspapers	PN	A to E	340	308,504
Yahoo! Answers	OC	A to B	938	93,932
Magazines	PM	A to D	86	202,268
Yahoo! Blog	OY	A to B	471	92,746

sion has any instances on the timeline or not. If an instance is recognised, the annotators define a ⟨MAKEINSTANCE⟩ in the corpus. The ⟨MAKEINSTANCE⟩ is a stand-off from the event expression, but is linked on the ⟨EVENT⟩ tag by the @eid attribute. Third, the annotators annotate the @class attribute on the ⟨EVENT⟩. @class attributes are nine: seven for event instances (OCCURRENCE, REPORTING, PERCEPTION, ASPECTUAL, I_ACTION, I_STATE, STATE) and two for non-instances (NULL and NONE). The difference between NULL and NONE is that the former is applied by ⟨EVENT⟩ annotators and the latter by ⟨TLINK⟩ annotators below. The instances are doubly checked by both ⟨EVENT⟩ and ⟨TLINK⟩ annotators. [2]

- OCCURRENCE: Event expressions without event arguments describing something that happens or occurs in the world (the argument event). Most event expressions belong to this class.
- REPORTING: Event expressions with an event argument describing the action of an animate actor declaring, narrating, or informing about the argument event.
- PERCEPTION: Event expressions with an event argument describing the physical perception of the argument event.
- ASPECTUAL: Event expressions with an event argument describing some aspectuality of the argument event.
- I_ACTION: Intensional action expressions with an event argument describing an action or situation to introduce the argument event, from which we can infer something given its relation with the I_ACTION.
- I_STATE: Intensional state expressions with an event argument referring to an alternative or possible world.
- STATE: State expressions in the timeline. We only annotate when an instance is introduced and becomes an argument of the other event expressions.
- NULL, NONE: Non-instance expressions.

The annotator discriminates whether the target is an event or a state (STATE). Then, he or she judges whether the target has any event argument or not (OCCURRENCE). Finally, he or she categorises any target with an event argument into one of the five categories of REPORTING, PERCEPTION, ASPECTUAL, I_ACTION, and I_STATE.

The two annotators and two supervisors defined a detailed linguistic annotation specification employing some Japanese language tests based on linguistic research (Kudo, 1995; Kudo, 2004; Nakamura, 2001). The two annotators are trained by the specification until the agreement rate reaches 75%.

3.3 ⟨TLINK⟩

⟨TLINK⟩ defines the temporal ordering of temporal information expressions and event expressions. We use a variant of Allen's interval algebra as ⟨TLINK⟩ labels; there are 13 labels for temporal ordering and three for event-subevent relations. We also have one label 'vague' for underspecified relations. Figure 2 shows the 13 + 3 labels. The three underlined labels — 'is_included', 'identity', and 'include' — are event-subevent versions of 'during', 'equal' and 'contains', respectively. Strictly, we can also define event-subevent

[2]Though the original TimeML defines 7 @class attributes on non instance event expressions, we do not define it. This is because our main research objective is annotating temporal ordering on the event instances.

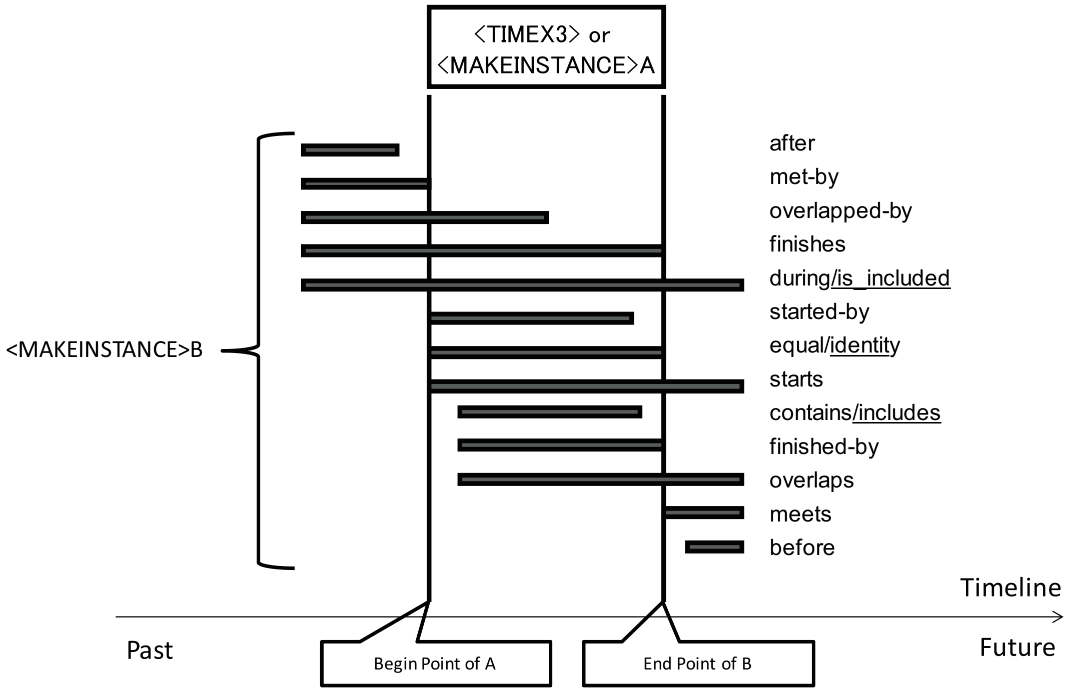

Figure 2: ⟨TLINK⟩ : Labels

versions of 'finishes', 'started-by', 'starts', and 'finished-by'. However, we did not define these, because they are rare and because TimeML did not define them.

⟨TLINK⟩ annotators are different from ⟨EVENT⟩ and ⟨MAKEINSTANCE⟩ annotators. Three annotators annotate the ⟨TLINK⟩ labels on part of pairs among ⟨TIMEX3⟩ and ⟨MAKEINSTANCE⟩. The number of ⟨TLINK⟩ candidates are square of the number of ⟨TIMEX3⟩ and ⟨MAKEINSTANCE⟩. It is hard to check all possible pairs in the documents; therefore, we limit the target pairs to the following four types of relations:

- 'DCT': relations between a ⟨TIMEX3⟩ of document creation time (DCT) and an event instance;
- 'T2E': relations between a ⟨TIMEX3⟩ (non DCT) and an event instance within one sentence;
- 'E2E': relations between two consecutive event instances; and
- 'MAT': relations between two consecutive matrix verbs of event instances.

If the relation is between two different possible worlds, we use the label 'vague'. When we regard the 'vague' relations as disjoint links , the connected subgraph indicates the different possible worlds.

The value of ⟨TIMEX3⟩ is regarded not as a time point but as a time interval. The event instance of a punctual verb is regarded as a time point occurrence, whereas the other event instances are regarded as time interval occurrences.

3.4 Other Tags in the original ISO-TimeML

In the original TimeML, ⟨SIGNAL⟩, ⟨SLINK⟩ , and ⟨ALINK⟩ are defined. ⟨SIGNAL⟩ is used with some temporal prepositions and conjunctions in English, ⟨SLINK⟩ is used for subordination relations, and ⟨ALINK⟩ is used for non-constituent aspectual relations. Currently, we are not using these with the BCCWJ-TimeBank.

4 BCCWJ-TimeBank

This section presents basic statistics on BCCWJ-TimeBank, the Japanese corpus annotated for temporal information. We also consider the annotation environment of BCCWJ-TimeBank.

4.1 ⟨TIMEX3⟩

We use XML Editor oXygen [3] for ⟨TIMEX3⟩ annotation. We define DTD for BCCWJ-TimeBank.

[3] http://www.oxygenxml.com/

The DTD enables us to use the machine-aided (in terms of, XML validation, a completion mechanism, and so on) environment of oXygen. An annotator performs inline annotation on the original text corpus. We introduce a pair-programming-like method in which a display is shared by an annotator and supervisor. Though the method is stressful for both annotator and supervisor, the data becomes more consistent and annotation errors are reduced.

Table 2 shows annotation target samples for ⟨TIMEX3⟩. The column 'W/TIMEX' shows the number of samples or sentences which include at least one temporal information expression. Some samples in the registers OW (white paper), OC (Yahoo! Answers), and OY (Yahoo! Blogs) do not include any temporal information expressions.

Table 3 shows the basic statistics of ⟨TIMEX3⟩ annotations. The table shows the number of ⟨TIMEX3⟩ by @type and @definite and the relation of {@value and @valueFromSurface}. @type has four labels: DATE, TIME, DURATION, and SET. We exclude document creation time (DCT), which is given in corpus metadata, from the statistics. Then, we analyse the statistics on the basis of two perspectives. The first is whether @definite is 'true' or 'false', in other words, whether the temporal information expression is fully specified or under-specified. The former can be mapped on the timeline, while the latter cannot. The other perspective is whether @value and @valueFromSurface are identical ('=') or not ('≠'). The former have undergone some normalisation procedure from the annotators, while the latter have not.

A total of 5,297 temporal information expressions are annotated in the corpus. Of those, 1,639 (30%) are fully specified expressions without any normalisation procedures applied. 2,023 (37%) of that can be normalised by contextual information, and 1,875 (34%) cannot. The third group need more external information to be normalised.

In the 'DATE' expressions, most of the fully-specified expressions(@definite 'true'; 61%) have had manual normalisation performed (@value ≠ @valueFromSurface;50%). This fact shows that the normalisation procedure is important for temporal information processing. The normalisation includes conversion from Japanese to western calendar, conversion from 2-digit to 4-digit western calendar, and completion year (taken from document creation time).

Table 4: ⟨TIMEX3⟩ : Statistics in PN (A)

⟨TIMEX3⟩ @type	
DCT (DATE)	54
DATE	727
TIME	107
DURATION	291
SET	19
ALL	1,198

Table 5: ⟨EVENT⟩ : Statistics in PN(A)

⟨EVENT⟩@class	Count
non-instance	(1,129)
NULL	1,114
NONE	15
event instance w/o event arg	(2,352)
OCCURRENCE	2,352
event instance w/ event arg	(1,291)
REPORTING	126
PERCEPTION	27
ASPECTUAL	63
I_ACTION	880
I_STATE	195
state instance	(181)
STATE	181

In the 'TIME' expressions, most fully specified expressions have had manual normalisation performed. The normalisation includes completion date (from document creation time) and resolution of a.m./p.m. ambiguity.

In the 'DURATION' and 'SET' expressions, @definite 'true' means that the length of the temporal region on the timeline can be uniquely determined. When we map on the timeline, we need ⟨TLINK⟩ information with 'DATE' or 'TIME' expressions or event expressions.

Note that we reduce the annotation target samples of ⟨EVENT⟩ , ⟨MAKEINSTANCE⟩ , and ⟨TLINK⟩ PN register (A) — 54 samples. The reason is that only the PN (newspaper) sample has date-level document creation time information as metadata. Table 4 shows the statistics of ⟨TIMEX3⟩ in PN (A) samples.

Table 2: ⟨`TIMEX3`⟩ : Annotation target samples

Register		# of Samples			# of Sentences			# of Words
		ALL	W/ TIMEX		ALL	W/ TIMEX		ALL
OW	(A)	17	16	(94%)	1,439	405	(28%)	58,336
PB	(A)	25	25	(100%)	2,568	289	(11%)	57,929
PN	(A,B)	110	110	(100%)	5,582	1,562	(28%)	116,834
OC	(A)	518	250	(48%)	3,479	488	(14%)	60,086
PM	(A)	23	23	(100%)	3,066	413	(13%)	59,372
OY	(A)	257	198	(77%)	3,986	765	(19%)	63,459

Table 3: ⟨`TIMEX3`⟩ : `@type` × `@definite` × {`@value` , `@valueFromSurface`}

`@definite`	`true` (fully-specified)						`false` (under-specified)					
`@value` and `@valueFromSurface`	all		=		≠		all		=		≠	
DATE	2,214	(61%)	381	(10%)	1,833	(50%)	1,438	(39%)	1,275	(35%)	163	(4%)
TIME	188	(37%)	1	(0%)	187	(37%)	315	(63%)	239	(48%)	76	(15%)
DURATION	1,129	(92%)	1,128	(92%)	1	(0%)	99	(8%)	99	(8%)	0	
SET	131	(85%)	129	(84%)	2	(1%)	23	(15%)	22	(14%)	1	(1%)
ALL	3,662	(66%)	1,639	(30%)	2,023	(37%)	1,875	(34%)	1,635	(30%)	240	(4%)

4.2 ⟨`EVENT`⟩ and ⟨`MAKEINSTANCE`⟩

We annotate ⟨`EVENT`⟩ and ⟨`MAKEINSTANCE`⟩ tags only on PN register (A). Table 5 shows the statistics of ⟨`EVENT`⟩ tags by `@class`. Event instances by ⟨`MAKEINSTANCE`⟩ are defined on the last seven `@class` in the tables. The number of ⟨`MAKEINSTANCE`⟩ is 3,839.

4.3 ⟨`TLINK`⟩

The three annotators are independently trained for ⟨`TLINK`⟩ annotation. The annotation is performed on four types of relations: 'DCT', 'T2E', 'E2E', and 'MATRIX'.

Table 6 shows annotation agreement among the 13+3+1 labels by three annotations and relation types. The three ∩-connected numbers are the label counts by each of the three annotators. The right number after '=' is the agreed count.

The agreed counts for 'after', 'during', 'contains', and 'before' are higher than the others. These relations do not exhibit boundary matching between the two time intervals. The relation 'equal' is the most frequent of those that do include interval boundary matching. Other relations are infrequent and show low agreement count among the three annotators. These findings show that a judgment of interval boundary matching is rare among and difficult for human annotators.

The relation 'vague' was agreed on 314 times by the three annotators. This fact shows that the discrimination of possible worlds might be doable by annotation.

Table 7 shows agreement rates by relation type across the three evaluation schemata. We define the schemata as follows: 'Label 13+3+1' is the most fine-grained evaluation schema; in it, all 13+3+1 relations are discriminated. 'Label 13+1' is a schema without event-subevent discrimination, in which 'is_included', 'identity', and 'includes' are regarded in the same light as 'during', 'equals', and 'contains', respectively. 'Label 5+1' is a TempEval-like schema in which 13+3+1 relations are generalised into 5+1 relations: 'BEFORE', 'BEFORE-OR-OVERLAP', 'OVERLAP', 'OVERLAP-OR-AFTER', 'AFTER', and 'VAGUE'.

The agreement rate across all relations is 65.3% (Cohen's kappa 0.733) using the most fine-grained evaluation schema (Label 13+3+1). We perform ⟨`TLINK`⟩ annotations on fixed relation pairs of four types. TimeBank 1.2 jointly performs ⟨`TLINK`⟩ annotations without fixing relation pairs. In this method, the ⟨`TLINK`⟩ relation agreement rate is 77% and the relation pairs agreement 55%. We believe that the BCCWJ-TimeBank ⟨`TLINK`⟩ relation agreement rate is in no way inferior to that of TimeBank 1.2. Among the four relation types, the agreement rate of 'DCT' is the highest and that of 'T2E' second-highest. The relation between a temporal information expression and an event instance is easier than the relation between two event instances. This is because the interval of

Table 6: ⟨TLINK⟩ : Annotation agreement by Annotator × Label × Relation type

Relation types	DCT	T2E	E2E	MATRIX	All
Count	3,839	2,188	2,972	1,245	10,244
after	2,352∩2,326∩2,133=1,961	396∩441∩432=315	627∩631∩639=432	292∩284∩277=198	3,667∩3,682∩3,481=2,906
met-by	0∩0∩0=0	5∩10∩2=2	18∩12∩3=2	7∩3∩2=1	30∩25∩7=5
overlapped-by	11∩5∩4=2	59∩52∩42=20	3∩3∩2=0	0∩0∩1=0	73∩60∩49=22
finishes	2∩8∩1=0	10∩1∩11=0	5∩8∩5=1	1∩0∩0=0	18∩17∩17=1
during	449∩424∩650=217	105∩100∩113=62	206∩139∩225=67	112∩86∩134=43	872∩749∩1,122=389
started-by	1∩0∩0=0	9∩2∩8=0	3∩14∩6=2	0∩3∩0=0	13∩19∩14=2
equal	1∩17∩0=0	37∩70∩51=19	263∩412∩307=154	62∩140∩90=29	363∩639∩448=202
starts	2∩0∩0=0	30∩9∩14=2	6∩16∩2=0	0∩1∩1=0	38∩26∩17=2
contains	164∩85∩144=63	830∩853∩868=671	299∩292∩344=117	148∩152∩188=64	1,441∩1,382∩1,544=915
finished-by	0∩0∩0=0	3∩3∩0=0	6∩7∩6=0	1∩3∩0=0	10∩13∩6=0
overlaps	2∩2∩4=1	75∩84∩70=32	6∩27∩5=0	1∩4∩3=0	84∩117∩82=33
meets	1∩13∩0=0	25∩26∩2=2	88∩88∩32=22	9∩15∩0=0	123∩142∩34=24
before	739∩767∩746=572	389∩360∩383=288	1,058∩994∩1,098=713	418∩436∩422=294	2,604∩2,557∩2,649=1,867
is_included	0∩0∩0=0	0∩0∩0=0	19∩2∩6=1	6∩0∩1=0	25∩2∩7=1
identity	0∩0∩0=0	0∩0∩1=0	11∩7∩24=2	16∩5∩15=2	27∩12∩40=4
includes	0∩0∩0=0	0∩0∩0=0	27∩10∩2=1	18∩2∩0=0	45∩12∩2=1
vague	115∩191∩157=38	212∩177∩191=100	327∩309∩265=128	154∩111∩111=48	808∩788∩724=314

Annotator A ∩ Annotator B ∩ Annotator C = Agreed count

Table 7: ⟨TLINK⟩ : Annotation agreement by Relation type across three evaluation schemata

Relation types	DCT	T2E	E2E	MATRIX	ALL
Count	3,839	2,188	2,972	1,245	10,244
Label 13+3+1	0.743(2854)	0.691(1513)	0.552(1642)	0.545(679)	0.653(6688)
Label 13+1	0.743(2854)	0.691(1513)	0.561(1667)	0.560(697)	0.657(6731)
Label 5+1	0.748(2873)	0.734(1605)	0.627(1862)	0.623(776)	0.695(7116)

Agreement rate(Agreed count)

the temporal information expression is more easily defined on the timeline than the interval of the event instance is. Under the relaxed relation evaluation schema, the agreement rates of 'E2E' and 'MATRIX' increase. This means that while interval boundary matching in these event instances is hard for the annotators to agree upon, interval anteroposterior relations can be agreed on.

Finally, table 8 shows agreement by two entity types: DCT and TIMEX of ⟨EVENT⟩@class. Relations with STATE tend to show low agreement rates, and relations between DCT/TIMEX and STATE are lower than relations between DCT/TIMEX and other event instances. This is because recognition of the time interval boundaries of state expressions is difficult for the annotators. In event instances with event arguments, relations with REPORTING, I_ACTION tend to show high agreement rates than averages.

5 Conclusions

This paper presents a temporal information-annotated Japanese specification and corpus. We adapt the temporal information annotation specification of the original TimeML and ISO-TimeML to the Japanese languages in several layers: ⟨TIMEX3⟩, ⟨EVENT⟩, ⟨MAKEINSTANCE⟩, and ⟨TLINK⟩. We construct BCCWJ-TimeBank as the realisation of the specification. Achieved temporal ordering agreement rates are 65.3%.

As ongoing research, we will continue to look into machine-learning-based temporal ordering estimation. In English temporal ordering, the tense and aspect information in ⟨MAKEINSTANCE⟩ are important features. However, in Japanese temporal ordering, the morphologically overt information is 'る (-ru)' vs 'た (-ta)' for non-past and past tense and 'る (-ru)' vs 'ている (-teiru)' for limited aspect. We will report the results of this temporal ordering estimation.

Further, as future research, we intend to take advantage of BCCWJ's status as the first balanced large-scale shared corpus of Japanese and analyse our annotation as compared to the syntactic and semantic annotations conducted on BCCWJ by several Japanese research institutes, as mentioned in section 2.2. Since Japanese is a modality-rich language, the modality annotations by these other institutes will be important for temporal ordering.

Acknowledgements

References

Y. Den, J. Nakamura, T. Ogiso, and H. Ogura. 2008. A proper approach to Japanese morphological analysis: Dictionary, model, and evalua-

Table 8: ⟨TLINK⟩ : Annotation agreement by {DCT, ⟨TIMEX3⟩ , ⟨EVENT⟩@class} × ⟨EVENT⟩@class

	DCT	TIMEX	OCC	REP	PER	ASP	I_A	I_S	STA	ALL
OCCURRENCE	0.739	0.702	0.551	0.625	0.286	0.718	0.559	0.592	0.422	0.656
Abbr. OCC	(2,352)	(1,320)	(1,602)	(104)	(7)	(39)	(494)	(130)	(102)	(6,159)
REPORTING	0.881	0.697	0.663	0.222	1.000	0.667	0.519	0.368	0.500	0.694
Abbr. REP	(126)	(66)	(95)	(9)	(2)	(3)	(52)	(19)	(12)	(385)
PERCEPTION	0.815	0.700	0.444	NaN	0.000	NaN	0.500	1.000	0.000	0.646
Abbr. PER	(27)	(10)	(18)	(0)	(1)	(0)	(6)	(1)	(1)	(65)
ASPECTUAL	0.714	0.615	0.545	1.000	0.000	0.000	0.643	0.000	0.000	0.627
Abbr. ASP	(63)	(52)	(44)	(6)	(2)	(2)	(14)	(1)	(1)	(185)
I_ACTION	0.808	0.720	0.576	0.690	0.667	0.765	0.631	0.527	0.333	0.698
Abbr. I_A	(880)	(567)	(491)	(29)	(6)	(17)	(309)	(55)	(51)	(2407)
I_STATE	0.651	0.686	0.490	0.250	0.750	0.429	0.545	0.875	0.333	0.594
Abbr. I_S	(195)	(86)	(145)	(4)	(4)	(7)	(55)	(16)	(15)	(527)
STATE	0.492	0.398	0.356	0.600	1.000	0.444	0.431	0.333	0.238	0.424
Abbr. STA	(181)	(83)	(118)	(5)	(3)	(9)	(51)	(9)	(21)	(481)
ALL	0.743	0.691	0.548	0.618	0.560	0.649	0.573	0.562	0.374	0.653
	(3,839)	(2,188)	(2,524)	(157)	(25)	(77)	(984)	(233)	(203)	(10,244)

Agreement rate (Agreed count)

tion. In *the 6th Language Resources and Evaluation Conference(LREC-2008)*, pages 1019–1024.

ISO. 2008. Iso dis 24617-1: 2008 language resource management – semantic annotation framework - part 1: Time and events. International Organization for Standardization.

M. Kudo. 1995. *Asupekuto, Tensu-Taikei to Tekusuto — Gendai Nihongo-no Jikan-no Hyougen*. Hituzi Syobo.

M. Kudo. 2004. *Nihongo-no Asupekuto, Tensu, Muudo-Taikei Hyoujun-go Kenkyu-wo Koete (in Japanese)*. Hituzi Syobo.

K. Maekawa. 2008. Balanced Corpus of Contemporary Written Japanese. In *Proc. of the 6th Workshop on Asian Language Resources (ALR-8)*, pages 101–102.

C. Nakamura. 2001. *Nihongo-no Jikan Hyougen*. Kuroshio Shuppan.

J. Pustejovsky, J. Casta no, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, and G. Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, pages 337–353.

J. Pustejovsky, K. Lee, H. Bunt, and L. Romary. 2010. ISO-TimeML: an International Standard for Semantic Annotation. In *Proceedings of LREC-2010*, pages 394–397.

N. UzZaman, H. Llorens, L. Derczynski, J. Allen, M. Verhagen, and J. Pustejovsky. 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, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

M. Verhagen, R. Saurí, T. Caselli, and J. Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, pages 57–62.