

# NAIST Japanese Timex Corpus

## 日本語時間表現・事象表現タグつきコーパスの作成

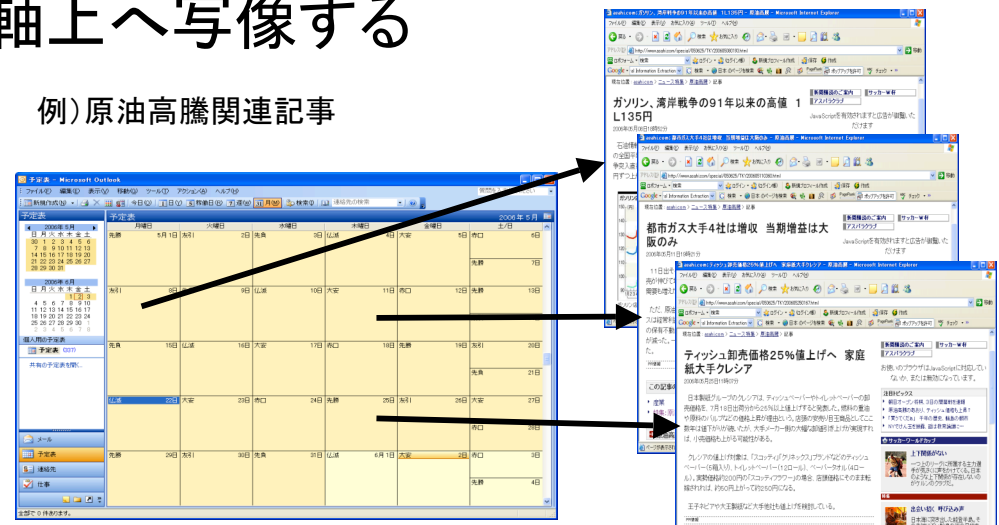
奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学講座  
浅原正幸 masayu-a@is.naist.jp

### 動機:

- 時間表現解析の精緻化  
→ 表現(mention)の切り出しだけでなく正規化(実時間軸上への写像)
- 事象表現解析システムの構築  
→ 事象が生起した時間を実時間軸上へ写像する

「RSS配信記事のカレンダーへの対応づけ」から  
「RSS配信記事中の出来事のカレンダーへの対応づけ」

例) 原油高騰関連記事



### 研究目的:

時間表現解析・事象表現解析  
を実現するための基礎データ  
としてタグつきコーパスを  
構築する。

記事配信日	出来事が起きる日	出来事
2006年5月8日	2006年5月8日	ガソリン最高値
2006年5月11日	2006年1月～3月	都市ガス大手4社増収
2006年5月25日	2006年7月18日	ティッシュ値上げ

# 付与する情報

TIDES/TimeML [Pustejovsky 02]/ISO8601に準拠  
カバーできない部分を独自に拡張

- ・時間表現に対するアノテーション

  - 〈TIMEX3/〉により表現

- ・事象表現に対するアノテーション

  - 〈EVENT/〉により各事象表現を切り出し、クラスを付与

  - 〈MAKEINSTANCE/〉により、各事象表現の文脈情報(ル形、テイル形、夕形、テイタ形の有無)を付与

- ・手がかり句に対するアノテーション

  - 〈SIGNAL/〉により表現

- ・時間表現⇔事象表現間、事象表現⇔事象表現間に対するアノテーション

  - 〈TLINK/〉により、表現間の順序関係を付与

# 時間表現に対するアノテーション(1/4)

## <TIMEX3/> タグにより表現属性に各種情報を付与する

### 属性名

·tid (#REQUIRED)

テキスト中の時間表現を表す識別子

·type (#REQUIRED)

(DATE|TIME|DURATION|SET)

DATE: 暦日表現

TIME: 時刻表現

DURATION: 期間を表す表現

SET: 時間集合を表す表現

·value (#REQUIRED)

時間表現が含意する値(推論して求める)

·type="DATE"

暦日表現に対して付与する。時刻まで指定されている場合には、次の“TIME”を用いる。

(例)「水曜日」「1980年」「7月18日」「昨日」「来週」「今冬」

·type="TIME"

時刻表現に対して付与する。

(例)「10時」「正午すぎ」「夜半」「未明」「今朝」「午前」「今」「昨夜中」

·type="DURATION"

具体的な時間が表層からはわからない、始点と終点がある時間表現に対して付与

する。

(例)「2ヶ月」「24時間」「3週間」「5分」「40歳」「30代前半」

·type="SET" 頻度などを表わす時間の集合に対して付与する。

(例)「週2回」「毎晩」「毎年」「3日おき」

Table 1: type="DATE" に対する value

単位	記号	時間表現例	value
年月日	XXXX-XX-XX	1980年7月7日	1980-07-07
曜日	XXXX-WXX-X	水曜日	XXXX-WXX-3
季節	XXXX-{SP,SU,FA,WI}	冬	XXXX-WI
四半期	XXXX-QX	第一四半期	XXXX-Q1
年度	FYXXXX	1998年度	FY1998
世紀	XXXX	11世紀	10XX
紀元前	BCXXXX	紀元前202年	BC0202
		4000年前	KA4
		2億年前	MA200

Table 4: type="TIME" に対する value

単位	記号	時間表現例	value
時刻	XXXX-XX-XXTXX:XX:XX	2006年8月8日午前8時45分30秒	2006-08-08T08:45:30
時刻(略記)	TXX:XX:XX	午前8時45分30秒	T08:45:30
その他	XXXX-XX-XXTXX	朝	XXXX-XX-XXTMO
		昼	XXXX-XX-XXTMI
		午後	XXXX-XX-XXTAF
		夕方	XXXX-XX-XXTEV
		夜	XXXX-XX-XXTNI
		日中	XXXX-XX-XXTDT

# 時間表現に対するアノテーション(2/4)

## value: 時間表現が含意する値

- value (#REQUIRED)  
時間表現が含意する値(推論して求める)
- valueFromSurface (#IMPLIED)  
時間表現の表層文字列から推定できる値
- temporalFunction (#IMPLIED)  
(false|true)  
value の推定に文脈情報が必要か否か
- valueFromFunction (#IMPLIED)  
時間表現の解析器から推定された文脈情報  
(現在未使用)

### ·Fully-specified vs. underspecified time expression

fully-specified time expression の例

“1998年9月10日”

“平成19年9月29日”

“2時間”

“週3回”

underspecified time expression の例

“9月10日”(年がわからない)

“昨年度”(現在時間を知る必要がある)

“9時20分”(日付がわからない)

underspecified time expression の場合、  
文脈情報からの推定が必要

temporalFunction=true

valueFromSurface ≠ value

### 2.4.5 不定な表現 (TIDES 4.3.2)

- value='PRESENT\_REF'  
現在を表わす不定な表現に用いる。  
(例)「今」「今日(こんにち)」「現在」  
  
<TIMEX3 type="DATE" value="PRESENT\_REF">今</TIMEX3>
- value='PAST\_REF'  
過去を表わす不定な表現に用いる。  
(例)「近年」「最近」「かつて」「過去」「むかしむかし」「以前」「ある日」「その日」
- value='FUTURE\_REF'  
未来を表わす不定な表現に用いる。  
(例)「未来」「明日(tomorrow 以外の意味)」「今後」「いつの日か」

Table 5: type="DURATION" に対する value

単位	記号	時間表現例	value
年	PnY	3年間	P3Y
月	PnM	2ヶ月	P2M
日	PnD	5日	P1D
時間	PTnH	3時間	PT3H
分	PTnM	30分	PT30M
秒	PTnS	9秒80	PT9.80S
週	PnW	1週間	P1W

# 時間表現に対するアノテーション(3/4)

## ·freq (#IMPLIED)

時間集合を表す表現に付与される頻度情報

## ·quant (#IMPLIED)

時間集合を表す表現に付与される量子子情報

- (例)

```
<TIMEX3 tid="t1" type="SET" value="P1D" quant="EACH">  
毎日  
</TIMEX3>  
ジョギングする。
```

- (例)

```
<TIMEX3 tid="t1" type="SET" value="P1W" freq="2X">  
週に2回  
</TIMEX3>  
ジョギングする。
```

## ·mod (#IMPLIED)

(BEFORE|AFTER|ON\_OR\_BEFORE|ON\_OR\_AFTER|LESS\_THAN|MORE\_THAN|EQUAL\_OR\_LESS|EQUAL\_OR\_MORE|START|MID|END|APPROX)

時間表現に対するモダリティ情報

### 2.6.1 時刻 (DATE, TIME) に対する mod 属性

- mod="BEFORE"  
(例) (2006年に記述されたとして)

```
<TIMEX3 type="DATE" value="1996-XX-XX" mod="BEFORE">  
10年以上前  
</TIMEX3>
```

### 2.4.4 type="SET"

- 「毎月」

```
<TIMEX3 type="SET" quant="each" value="P1M">毎月</TIMEX3>
```

- 「毎週」

```
<TIMEX3 type="SET" quant="each" value="P1W">毎週</TIMEX3>
```

### 2.6.2 時間 (DURATION) に対する mod 属性

- mod="EQUAL\_OR\_LESS"  
(例)

```
<TIMEX3 type="DURATION" value="PT10M" mod="EQUAL_OR_LESS">  
10分以内  
</TIMEX3>
```

- mod="EQUAL\_OR\_MORE" (例)

```
<TIMEX3 mod="EQUAL_OR_MORE" type="DURATION" value="PT1H">  
1時間以上  
</TIMEX3>
```

### 2.6.3 時刻 (DATE, TIME)、時間 (DURATION) に対する mod 属性

- mod="START"  
(例)

```
<TIMEX3 type="DATE" value="196X-XX-XX" mod="START">  
60年代初頭  
</TIMEX3>
```

- mod="MID"  
(例)

```
<TIMEX3 type="DATE" value="XXXX-10-XX" mod="MID">  
10月半ば  
</TIMEX3>
```

# 時間表現に対するアノテーション(4/4)

## •anchorTimeID(#IMPLIED)

時刻表現に対する共参照情報

## •beginPoint

時間表現に対する共参照情報(開始時刻)

## •endPoint

時間表現に対する共参照情報(終了時刻)

## •functionInDocument(#IMPLIED)

(CREATION\_TIME|EXPIRATION\_TIME|MODIFICATION\_TIME|PUBLICATION\_TIME|RELEASE\_TIME|RECEPTION\_TIME|NONE)

文書作成日時などの情報(現在は CREATION\_TIME=Document Creation Time のみ)

## •comment(#IMPLIED)

コメント(自由記述形式)

### • (例)

彼は  
<TIMEX3 tid="t1" type="DATE" value="1998-04-02">  
1998年4月2日  
</TIMEX3>  
に奈良にやってきた。  
<TIMEX3 tid="t2" type="DATE" value="1998-04-03" anchorTimeID="t1">  
翌日  
</TIMEX3>、  
彼ははじめて学校に行った。

外界照応の場合は特殊な timeID t0 を用いる。

### • (例)

彼は  
<TIMEX3 tid="t1" type="DATE" value="XXXX-WXX-4" anchorTimeID="t0">  
木曜  
</TIMEX3>  
奈良にやってきた。

### • (例)

彼は  
<TIMEX3 tid="t1" type="DATE" value="XXXX-09-15">  
9月15日  
</TIMEX3>  
から  
<TIMEX3 tid="t2" type="DATE" value="XXXX-09-22">  
9月22日  
</TIMEX3>  
まで  
<TIMEX3 tid="t3" type="DURATION" value="P1W" beginPoint="t1" endPoint="t2">  
1週間  
</TIMEX3>  
旅行にでかける。

# 事象表現に対するアノテーション

## <EVENT/>タグ

事象表現そのものに関する情報

属性名

- eid (#REQUIRED)  
テキスト中の事象表現を表す識別子
- class (#REQUIRED)  
事象表現に対するクラス付与

## <MAKEINSTANCE/>タグ

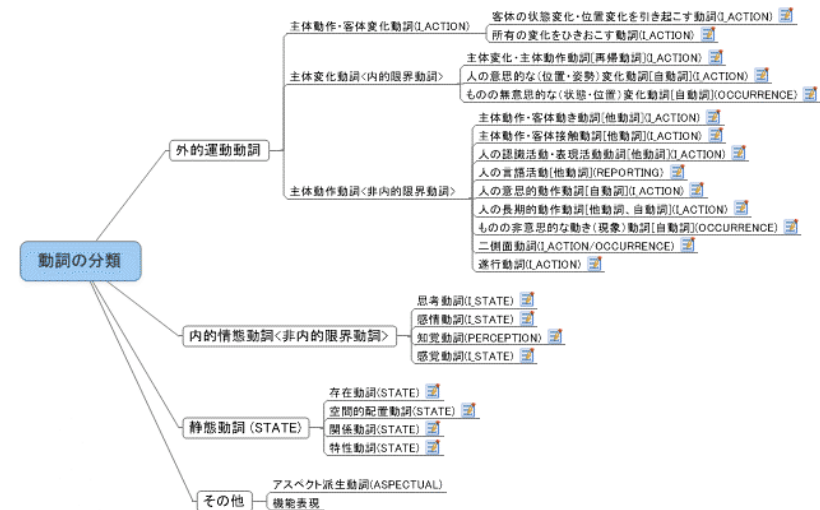
文脈に基づく事象表現に関する情報

属性名

- eiid (#REQUIRED)  
MAKEINSTANCE の識別子
- eventid (#REQUIRED)  
テキスト中の事象表現の eid を指す
- tense (#REQUIRED)  
文脈からのテンスに関する情報(PAST|PRESENT|FUTURE|NONE)
- aspect (#REQUIRED)  
文脈からのアスペクトに関する情報
- nf\_morph (#REQUIRED)  
不定表現に対する情報(ADJECTIVE|NOUN|NONE)
- polarity (#IMPLIED)  
待遇表現に関する情報
- modality (#IMPLIED)  
モダリティに関する情報
- signalID (#IMPLIED)  
関連する手がかり句
- cardinality (#IMPLIED)  
事象に対する頻度情報

現在、何を付与するか検討中

- TimeML オリジナルのクラス(下図参照)  
(REPORTING|PERCEPTION|ASPECTUAL|I\_ACTION|I\_STATE|STATE|OCCURENCE)
- 工藤らの動詞の分類(下図参照)



工藤らの動詞の分類(TimeML EVENT:class との対応)

•中村らの素性構造

point (時点)

inc-p (inchoative point, 始動点)

ter-p (terminative point, 終了点)

duration (期間)

pre-dur (preparatory duration, 準備期間)

pro-dur (process duration, 過程期間)

res-dur (resultative duration, 結果期間)

# 手がかり句に対するアノテーション

## <SIGNAL/>タグにより表現

複文などを表す際に用いる接続表現に付与  
「ので」「ため」「とき」「したとたんに」…

# 表現間の関係に対するアノテーション(1/2)

## <TLINK/>タグにより表現

属性名

・lid (#IMPLIED)

TLINK の識別子

・(timeID|eventID)

比べる対象の時間or事象表現のID

・(relatedToTime|relatedToEvent)

比べられる対象の時間or事象表現のID

・relType

2表現間の時間的順序関係(次のポスター)

・origin (#IMPLIED)

何によってタグづけられたか

・人手によるもの

・推移律により類推されたもの

・signalID

間にある手がかり句のID

各表現が生起した時点・時間の順序関係をタグづけする

時間表現 ←→ 事象表現

国連安保理は19日、決議を採択した。

19日 DURING\_INV 採択し

事象表現 ←→ 事象表現

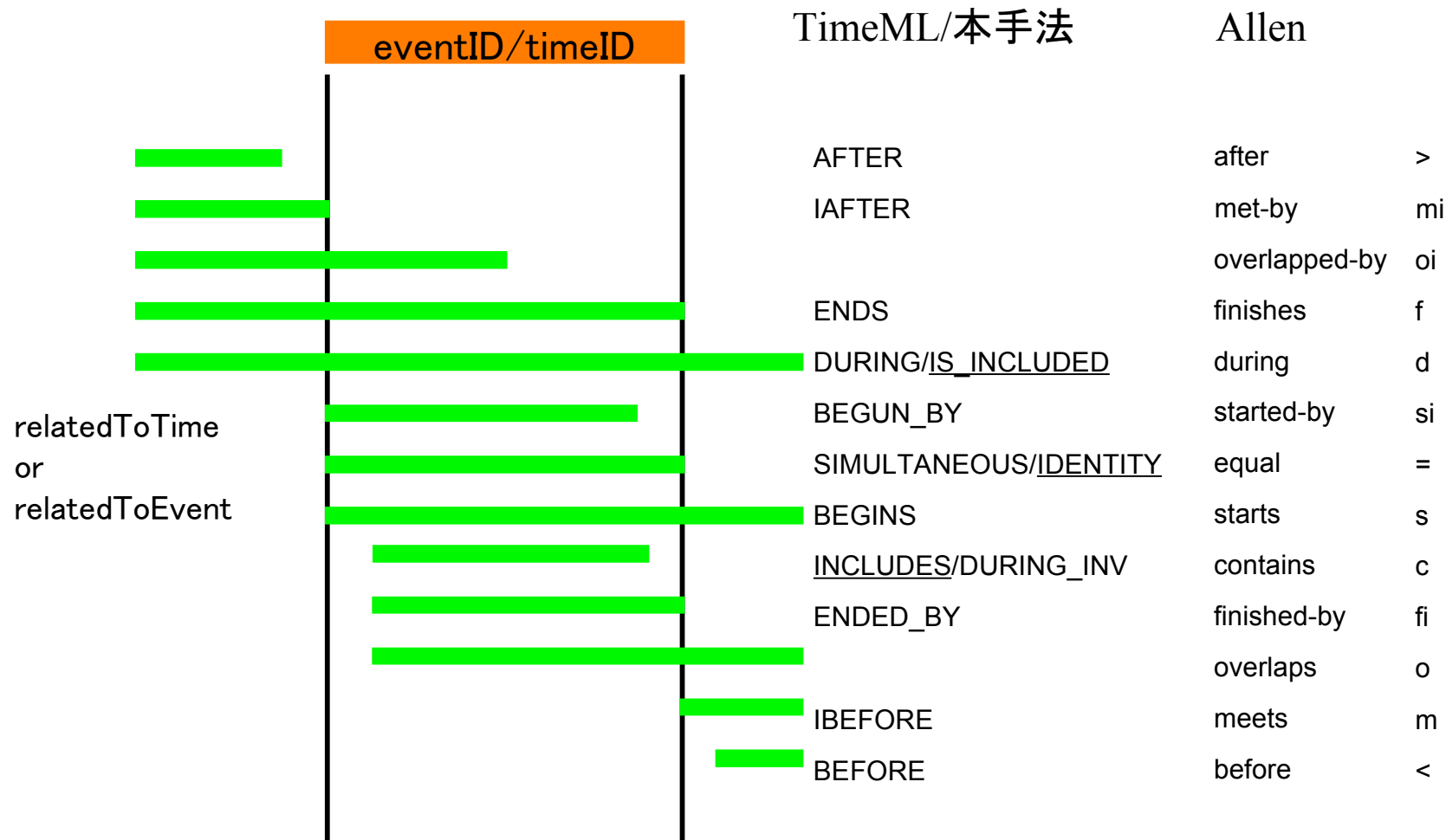
旅客機が山の衝突した瞬間、全乗客・乗員が即死した。

衝突し IBEFORE(immediately before) 即死し



# 表現間の関係に対するアノテーション(2/2)

TLINK@relType



注: IS\_INCLUDED/IDENTITY/INCLUDES は event-subevent 間の関係

例) “卵を割る” is\_included “オムライスを作る” / “車を買う” identity “車を購入する”

# 動詞句に関する知識獲得と事象の生起順序関係解析

- ・ VerbOcean(En) [Chklovski 2004]→[Pantel 2007]
  - ・ 動詞句間の関係 “HAPPENS-BEFORE”  
例)「むすぶ」→「ほどく」、「結婚する」→「離婚する」
  - ・ 動詞句間関係の選択選好モデルの構築
- ・ [Inui 2004] (Ja)
  - ・ タメ複文の主従節間からの動詞句間関係獲得
- ・ [Torisawa 2006](Ja)
  - ・ 時間軸上の制約(テ形接続)による動詞句間の関係獲得
- ・ [乾 2005](Ja)
  - ・ 因果関係タグつきコーパスの構築と分析
- ・ [Mani 2006]→[TempEval 2007] (En)
  - ・ 事象の生起順序関係を解析するために、VerbOceanを利用
- ・ [Cheng 2007](Zh)
  - ・ 依存構造解析結果に基づき限られた、事象対にのみ生起順序関係をタグづけ。推移閉包を作成し被覆率をあげる。