



Reasoning about Temporal and Event Information in Clinical Texts

James Pustejovsky

Brandeis University

National Institutes of Health

Bethesda, MD

April 23, 2012



Outline

- Why time is relevant but hard to understand
- Why standards are important to our field
- Properties of TimeML
- Increasing *informativeness* of annotation
- Application to Temporal Analysis in Clinical Texts



Motivation for time and event

- Natural language is filled with references to past and future events, as well as planned activities and goals;
- Without a robust ability to identify and temporally situate events of interest from language, the real importance of the information can be missed;
- A Robust Annotation standard can help leverage this information from natural language text.



Temporal Awareness in Real Text

- The bridge **collapsed** **during** the **storm** but **after** traffic was **rerouted** to the Bay Bridge.
- President Roosevelt **died** in April 1945 **before**
 - the war **ended**. (*event happened*)
 - he **dropped** the bomb. (*event didn't happen*)
- The CEO **plans** to **retire** **next** **month**.
- **Last** **week** Bill was **running** the marathon when he **twisted** his ankle. Someone had **tripped** him. He **fell** and didn't **finish** the race.



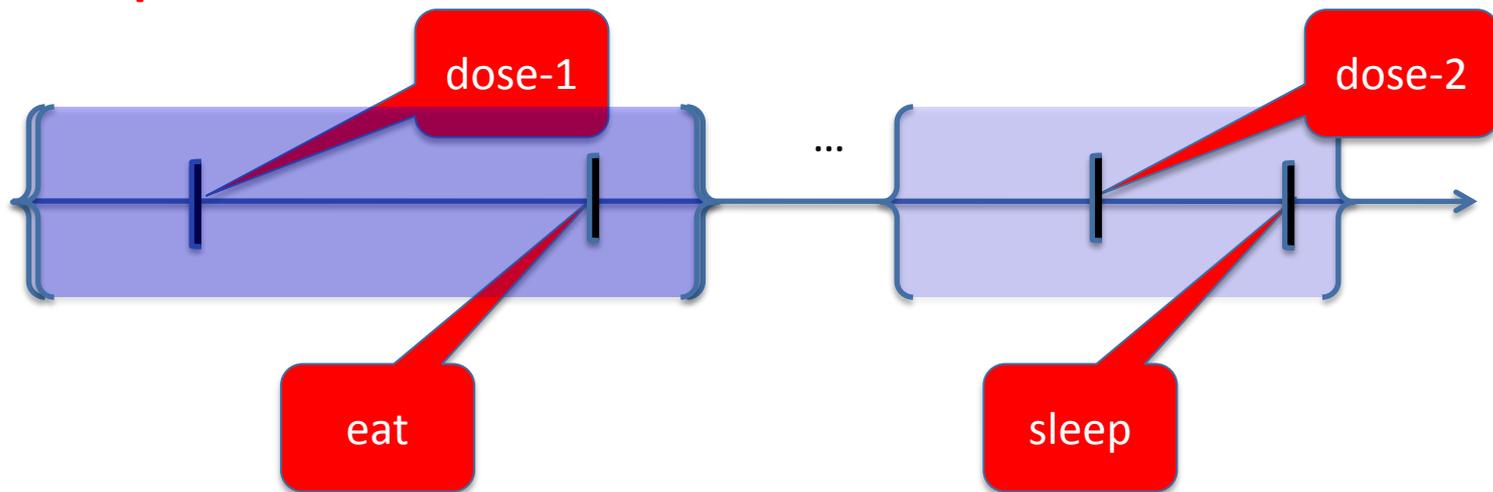
Temporal Awareness Challenge

- Identification of all important events in a text
- Actual temporal ordering and time anchoring of these events to temporal expressions.



Temporal Awareness

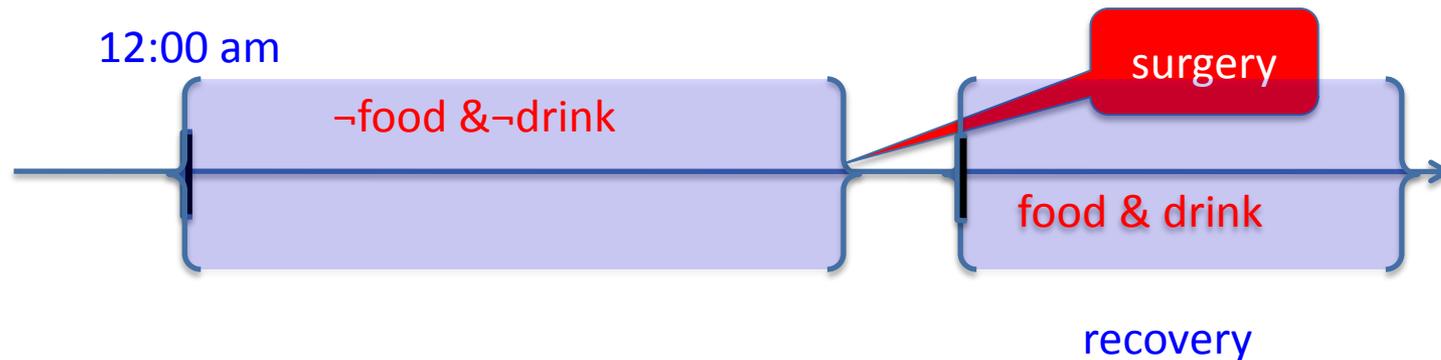
- Take your 1st **dose** of levaquin in the **morning** before any **food**, 2nd **dose** before **sleep**.





Temporal Awareness

- No **food** or **drink** after **midnight** before **surgery**, until you are in **recovery**.





Current Time Analysis Technology

- **Document Time Linking**
 - Find the document creation time and link that to all events in the text;
- **Local Time Stamping**
 - find an event and a “local temporal expression”, and link it to that time;



Time Stamping: the good, bad, ...

- ✓
• ☺ Set up a **meeting on Tuesday** with EMC.
- ✓
• ☺ Franklin **arrives tomorrow** from London.
- X
• ☹ Franklin **arrives on the afternoon** flight from London **tomorrow**.
- X
• ☹ ☹ Most people **drive today** while talking on the phone.



Document Time Stamping

April 25, 2010

- President Obama **paid tribute** Sunday to 29 workers **killed** in an explosion at a West Virginia coal mine earlier this month, **saying** they **died** "in pursuit of the American dream." The **blast** at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years. Obama **ordered** a **review** earlier this month and **blamed** mine officials for lax **regulation**.



Document Time Stamping:

April 25, 2010

- President Obama **paid tribute** Sunday to 29 workers **killed** in an explosion at a West Virginia coal mine earlier this month, **saying** they **died** "in pursuit of the American dream." The **blast** at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years. Obama **ordered** a **review** earlier this month and **blamed** mine officials for lax **regulation**.



Modeling Events relative to Time:

- **ORDER:**
 - The position of the interval relative to others :
- **MEASURE:**
 - The size of the interval;
- **QUANTITY:**
 - The number of intervals.



Identifying Temporal Relations

- A relation between two events (Discourse-constrained);
- A relation between two times (Domain-constrained);
- A relation between a time and an event (syntax and pragmatics of domain).



TimeML: Annotation of temporal entities

- **Temporal expressions: <TIMEX3>**
 - **Times:** *3 o'clock, mid-morning.*
 - **Dates:**
 - Fully Specified: *June 11, 1989; Summer, 2002.*
 - Underspecified: *Monday, next month, two days ago.*
 - **Durations:** *three months, two years.*
 - **Sets:** *every month, each Tuesday.*
- **Event expressions: <EVENT>**
 - **Expressions denoting events** that participate in the narrative of a given document, and which can be temporally ordered.
 - Event-related **grammatical features:**
 - **Tense:** past, present, past, etc.
 - **Aspect:** progressive, perfective, perfective-progressive.
 - **Polarity:** positive, negative.
 - **Modality:** *would, could, may,* etc.
 - **Class:** occurrence, state, aspectual, intensional, etc.



TimeML: Annotation of temporal relations

- **Temporal links: <TLINK>**
 - Anchoring of Events to Times
 - Ordering of Events
 - 13 temporal relations (based on Allen's relations), among which:
 - **Simultaneous**
 - **Before** (e.g., *For most of the **murders**, suspects have already been **arrested***)
 - **After**
 - **Immediately before** (e.g., *All passengers **died** when the plane **crashed** into the **mountain***)
 - **Immediately after.**
 - **Including** (e.g., *John **arrived** in Boston **last Thursday***)
 - Etc.
- **Aspectual links: <ALINK>** Phases of an event
 - **Initiation:** *John **started** to read.*
 - **Culmination:** *John **finished** assembling the table.*
 - **Termination:** *John **stopped** talking.*
 - **Continuation:** *John **kept** talking.*



TimeML: Annotation of temporal relations

- **Subordinating links <SLINK>**

- Events that syntactically subordinate other events
- Providing information about the factual nature of the embedded event:

Factive: The embedded event is presupposed or entailed as factual.

John forgot that he was in Boston last year.

Mary regrets that she didn't marry John.

Counterfactive: The embedded event is presupposed as non-factual:

John forgot to buy some wine.

John prevented the divorce.

Evidential: Introduced by REPORTING or PERCEPTION:

John said he bought some wine.

Mary saw John carrying only beer.

Negative evidential: Introduced by REPORTING events conveying negative polarity:

John denied he bought only beer.

Modal: Expressing different degrees of uncertainty, possibility, thought, etc.

Analysts also suspect suppliers have fallen victim to their own success.



Event Anchoring

- John taught on Tuesday.
 - a. EVENT tag introduces a quantified event expression $\implies \exists e_1[teach(e_1)]$;
 - b. TIMEX3 tag introduces the temporal expression $\implies \exists t_2[tuesday(t_2)]$;
 - c. TLINK introduces the ordering relation $\implies \lambda y \lambda x[\tau(x) \subseteq y]$.
-

a. $teach = e_1, tuesday = t_2$

b. $\exists e_1 \exists t_2 [teach(e_1) \wedge tuesday(t_2) \wedge \tau(e_1) \subseteq t_2]$



Event Ordering

- John taught before Mary arrived.

a. EVENT tag introduces a quantified event expression $\implies \exists e_1[teach(e_1)]$;

b. EVENT tag introduces a quantified event expression $\implies \exists e_2[arrive(e_2)]$;

c. TLINK introduces the ordering relation $\implies \lambda y \lambda x[\tau(x) < \tau(y)]$.

a. $teach = e_1, arrive = e_2$

b. $\exists e_1 \exists e_2[teach(e_1) \wedge arrive(e_2) \wedge \tau(e_1) < \tau(e_2)]$



Durations in TimeML

- John slept for two hours.

- Use TLINK:

```
<EVENT id="e1" pred="SLEEP"/>
```

```
<TIMEX3 id="t1" type="DURATION" value="P2H"/>
```

```
<TLINK eventID="e1" relatedToTime="t1" relType="SIMULTANEOUS"/>
```

a. $\text{sleep} = e_1, t_1 = \text{"two hours"}$

b. $\exists e_1 \exists t_1 [\text{sleep}(e_1) \wedge \text{two_hours}(t_1) \wedge \tau(e_1) = t_1]$



Measuring Events in ISO-TimeML

- John taught for three hours on Tuesday.
- Introduce MLINK:

```
<EVENT id="e1" pred="TEACH"/>
```

```
<TIMEX3 id="t2" type="DURATION" value="P3H"/>
```

```
<MLINK eventID="e1" relatedToTime="t2" />
```

a. $teach = e_1, tuesday = t_2, m = [3, hour]$

b. $\exists e_1 \exists t_2 [teach(e_1) \wedge \mu(\tau(e_1)) = m \wedge$
 $m = [3, hour] \wedge tuesday(t_2) \wedge \tau(e_1) \subseteq t_2]$



Specification Language (Semantics)

- Defines the expressiveness of representation
- Specifies the abstract syntax of the markup
- Identifies the semantics of the syntactic elements



Annotation Strategies (Pragmatics)

- How do we apply the specification to a text?
- Assuming we cannot be complete, what do we leave out?
- How do we compare annotations that are incomplete?



TimeML Guidelines

- The guideline does not specify certain types of annotations that *should* be performed;
- The guideline forces some annotations to be performed when they are *irrelevant*.



Document Creation Time

- 4-10-2011
- Local officials **reported yesterday** that a car **exploded** in downtown Basra.



TimeBank Annotation

- 4-10-2011
- Local officials **reported yesterday** that a car **exploded** in downtown Basra.
- DCT= t1, val=10-04-2011
- t2 = yesterday, val=09-04-2011
- e1 = report
- e2 = explode
- TLINK1 = before(e1; t1)
- TLINK2 = before(e2; t1)
- TLINK3 = includes(t2; e1)



Missing TLINKs

- 4-10-2011
- Local officials **reported yesterday** that a car **exploded** in downtown Basra.
- TLINK4 = includes(t2; e2)
- e2 = explode
- t2 = yesterday, val=09-04-2011



Style, Genre, Channel, and Anchors

- Channel: is the document written or spoken?
- Production circumstances: how was the document distributed? broadcast, newswire, daily publication;
- Style: what format was used to present the information?
- Presence of a temporal anchor: Whether an article contained a Narrative Time in the first sentence of the document.



Narrative Scope 1/2

- *Narrative Scope* describes the timespan described in the document, with the left marker defined by the earliest event mentioned in the document, and the right by the event furthest in the future.
- *Narrative Time* is essentially the current temporal anchor for events in a document, and can change as the reader moves through the narrative.



Narrative Scope

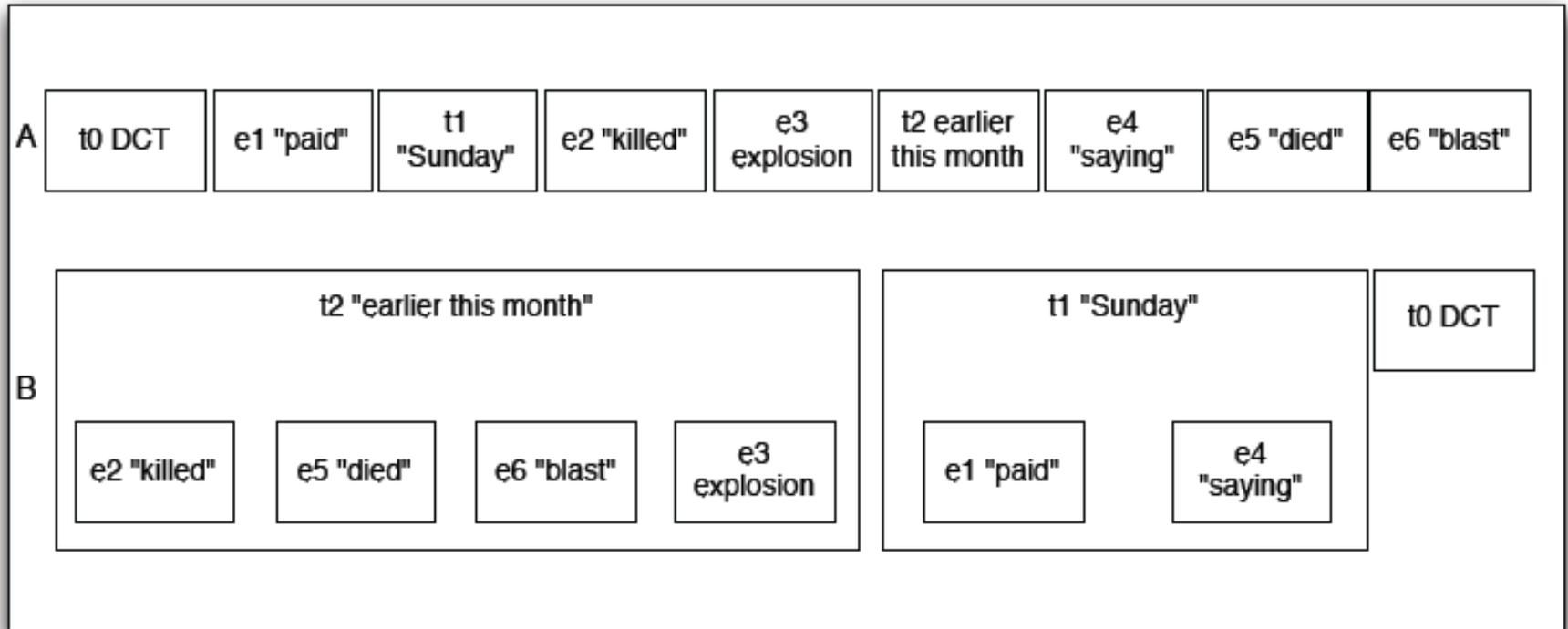
April 25, 2010 7:04 p.m. EDT -**t0**

S1: President Obama paid-**e1** tribute Sunday -**t1** to 29 workers killed-**e2** in an explosion -**e3** at a West Virginia coal mine earlier this month- **t2**, saying-**e4** they died-**e5** “in pursuit of the American dream.”

S2: The blast-**e6** at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.



Narrative Scope



A: Times and events as appearing in the text; B: events grouped into their appropriate Narrative Times.



Informativeness

- *External Informativeness*: information derived from relations outside the temporal relation constraint set, e.g., as coming from explicit discourse relations between events;
- *Internal Informativeness*: information derived from the nature of the relation itself, as defined largely by the algebra of relations.



Document Time Stamping: for real

April 25, 2010

- President Obama **paid tribute Sunday** to 29 workers **killed** in an **explosion** at a West Virginia coal mine **earlier this month**, **saying** they **died** "in pursuit of the American dream." The **blast** at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years. Obama **ordered** a **review earlier this month** and **blamed** mine officials for lax **regulation**.



Narrative Time

April 25, 2010

- President Obama **paid tribute Sunday** to 29 workers **killed** in an **explosion** at a West Virginia coal mine **earlier this month**, **saying** they **died** "in pursuit of the American dream." The **blast** at the Upper Big Branch Mine was the worst U.S. mine **disaster** in nearly 40 years. Obama **ordered** a **review earlier this month** and **blamed** mine officials for lax **regulation**.



Applying Temporal Processing to Clinical Texts

- **TimeText**: Zhou, Hripcsak, Parsons
- **CLEF**: Harkema et al.
- **ConText**: Chapman et al
- **THYME**: Savova and Palmer

- Let's explore the domain of clinical notes with TimeML in the THYME research:
 - Slides from Will Styler, Martha Palmer, Guergana Savova, Jim Martin, Wayne Ward, Jinho Choi

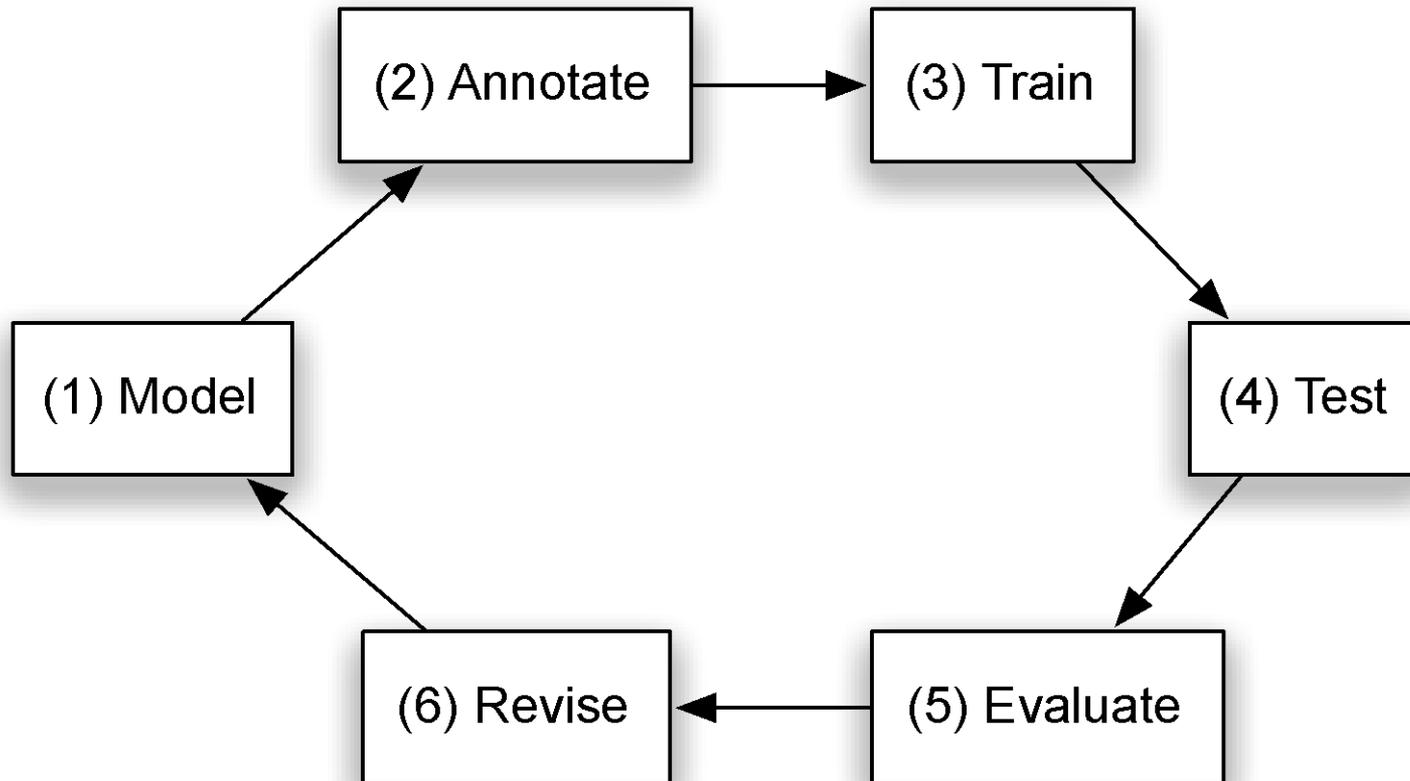


Semantic Requirements for Annotation

- Fundamental distinction between the concepts of *annotation* and *representation*
 - Based on ISO CD 24612 *Language resource management - Linguistic Annotation Framework* (Ide and Romary, 2004)
- Distinguish between *abstract syntax* and *concrete syntax*
 - Concrete Syntax → XML encoding
 - Abstract Syntax → Conceptual inventory and a set of syntactic rules defining the combination of these elements

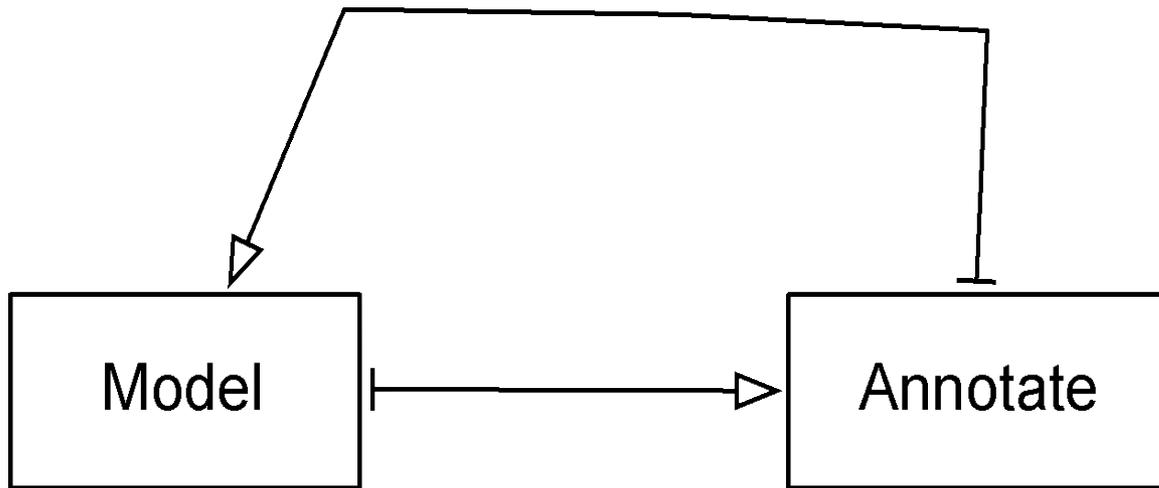


Specification Development Cycle: MATTER





Babbling Phase of Cycle





Entity Types

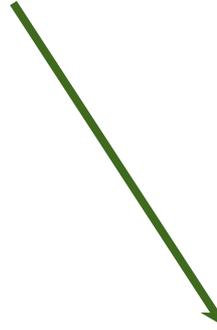
PERSON



DISORDER



PHENOMENON



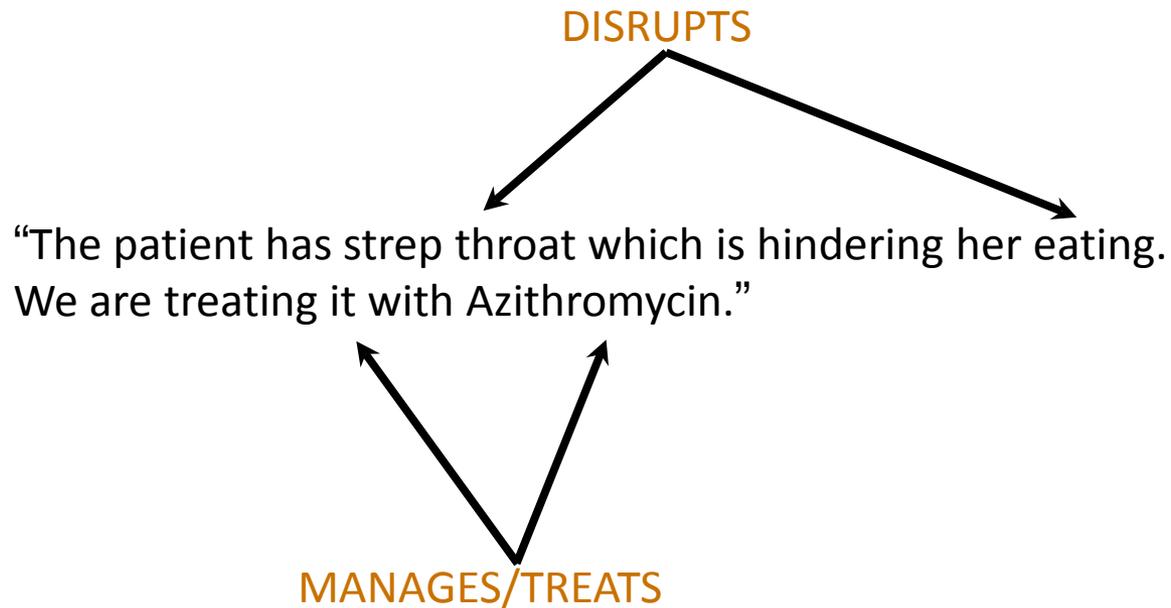
“The patient has strep throat which is hindering her eating. We are treating it with Azith



CHEMICAL/DRUG



Relations





Temporal Relation Annotation

- First, mark spans of text as either **EVENTs** or as Temporal Expressions (**TIMEX3s**) (like dates or times)
- Then, mark the relations between **EVENTs** and **TIMEX3s** using either Temporal Links (**TLINKs**) or Aspectual Links (**ALINKs**)



EVENT and TIMEX3

“A rash developed following surgery on 6/15, which we treated with hydrocortisone”

EVENTs:

rash (occurred before document time)

surgery (before document time)

treated (before document time)

hydrocortisone (before document time)

TIMEX3s:

6/15



TLINK

A rash developed following surgery on 6/15, which we treated with hydrocortisone.

- rash AFTER surgery
- treated AFTER rash
- treated AFTER surgery
- hydrocortisone AFTER surgery
- hydrocortisone OVERLAP treated
- 6/15 CONTAINS surgery
- rash AFTER 6/15
- hydrocortisone AFTER 6/15
- (and so forth)



ALINK

Links aspectual EVENTS to the EVENTS they modify

Four values:

CONTINUES, INITIATIVES, REINITIATES, and TERMINATES

“We will **continue** treating the **rash**”

Continue (**EVENT** of type Aspectual)

Treating (**EVENT** of type Process)

Rash (**EVENT** of type State)

ALINK: **continue** **CONTINUES** **treating**



Possible Temporal Inferences

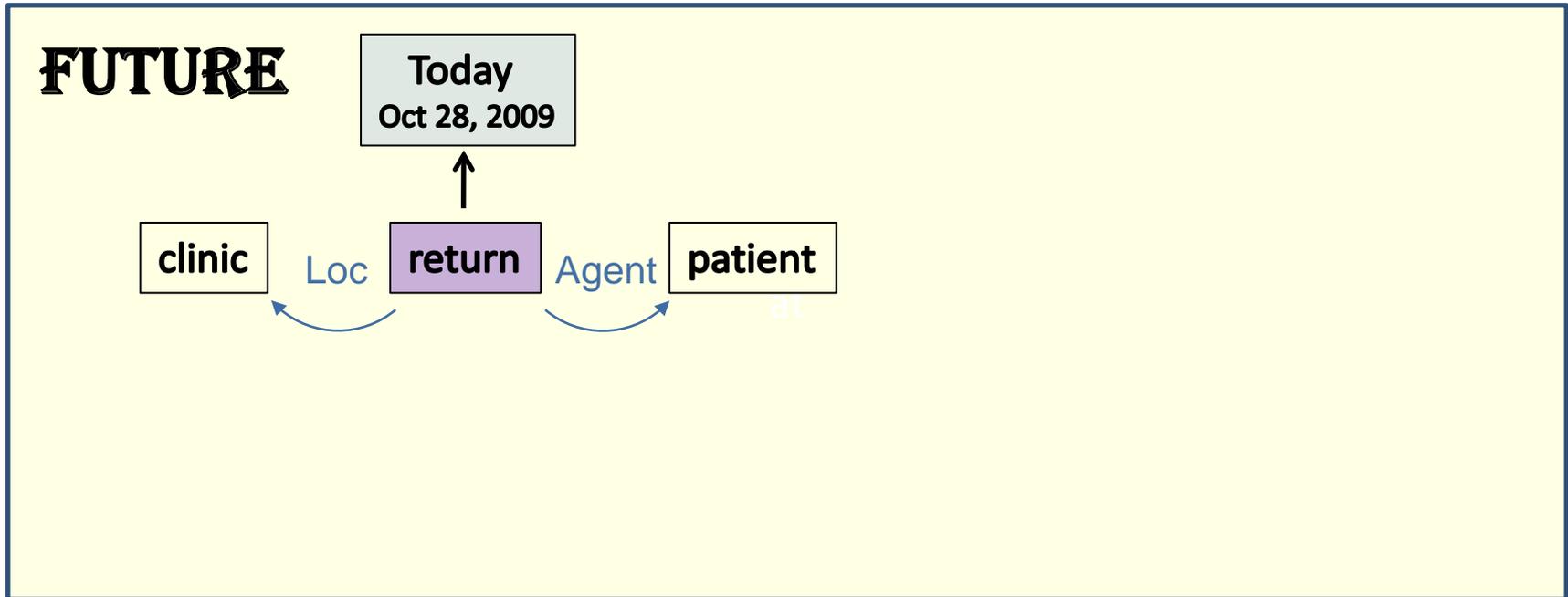
A rash developed following surgery on 6/15, which we treated with hydrocortisone.

- 1) The rash, surgery and treatment all occurred before this was written
- 2) The surgery occurred on June 15th
- 3) The rash developed after the surgery
- 4) The treatment occurred after the rash
- 5) The treatment occurred after the surgery



Integrating Time with Semantic Roles

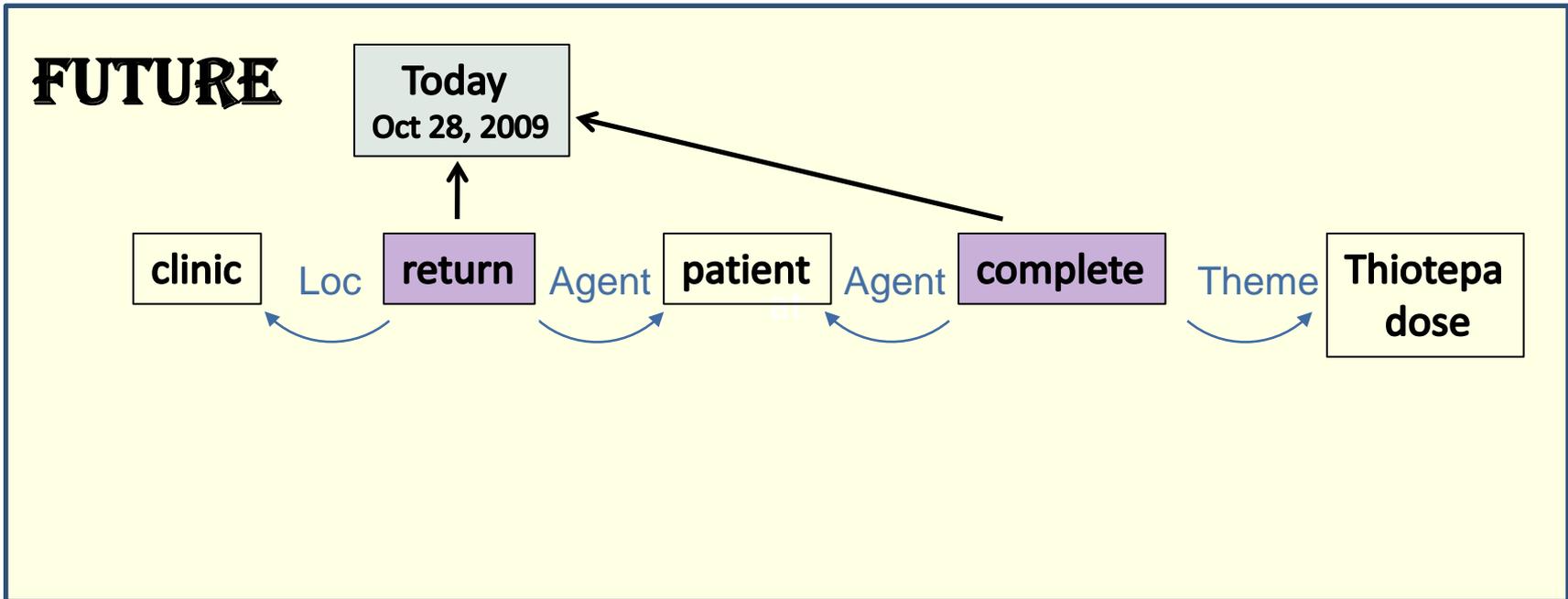
The patient returns to the outpatient clinic today for follow-up



the patient will complete his thiotepa dose today , and he will return tomorrow for the last dose of his thiotepa .
His donor completed stem-cell collection yesterday



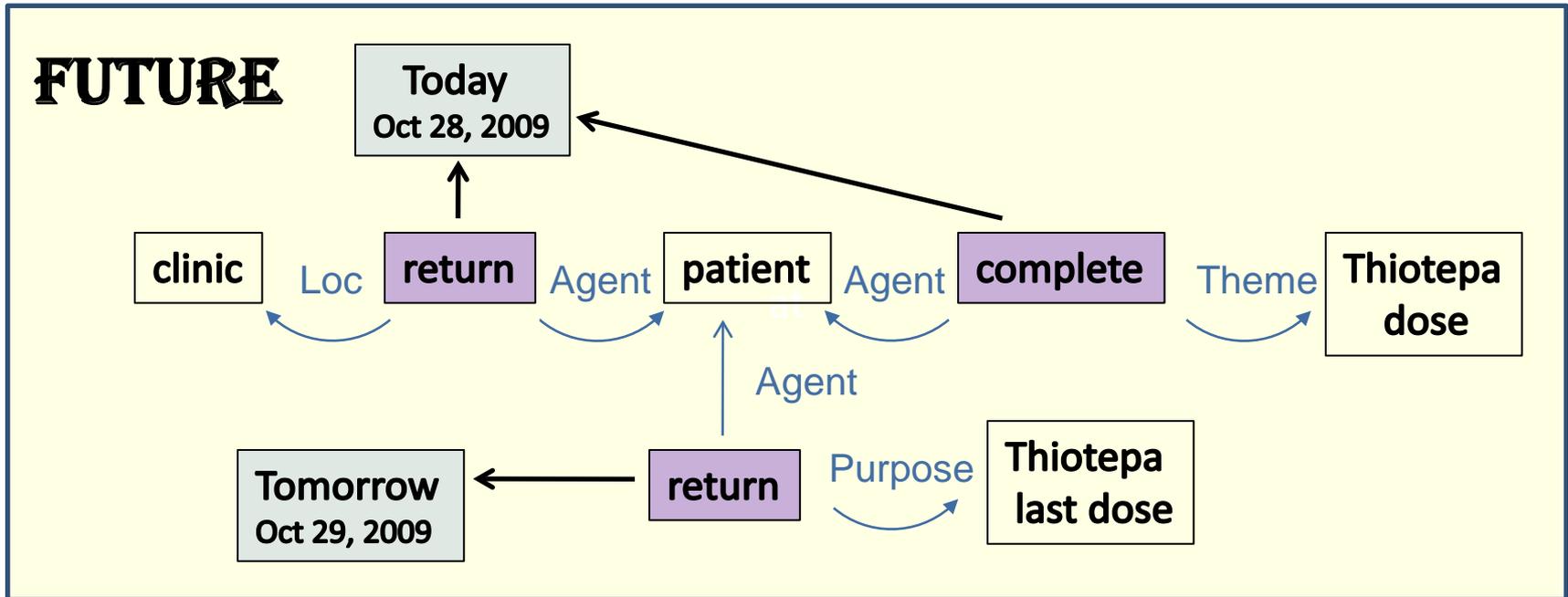
The patient returns to the outpatient clinic today for follow-up
the patient will complete his thiotepa dose today



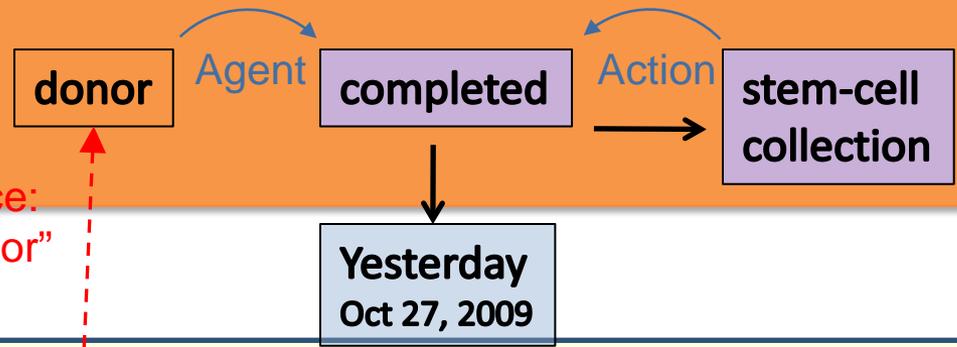
, and he will return
tomorrow for the last dose of his thiotepa .
His donor completed stem-cell collection yesterday



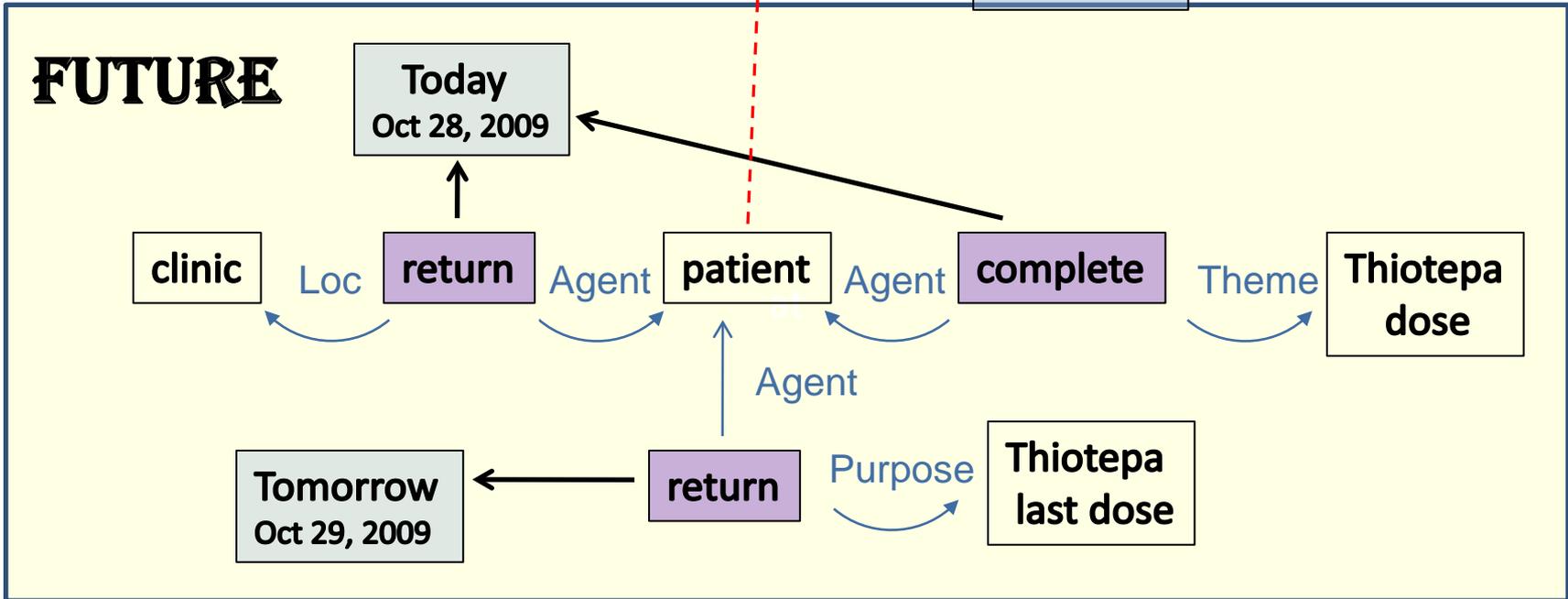
The patient returns to the outpatient clinic today for follow-up the patient will complete his thiotepa dose today , and he will return tomorrow for the last dose of his thiotepa .



His donor completed stem-cell collection yesterday

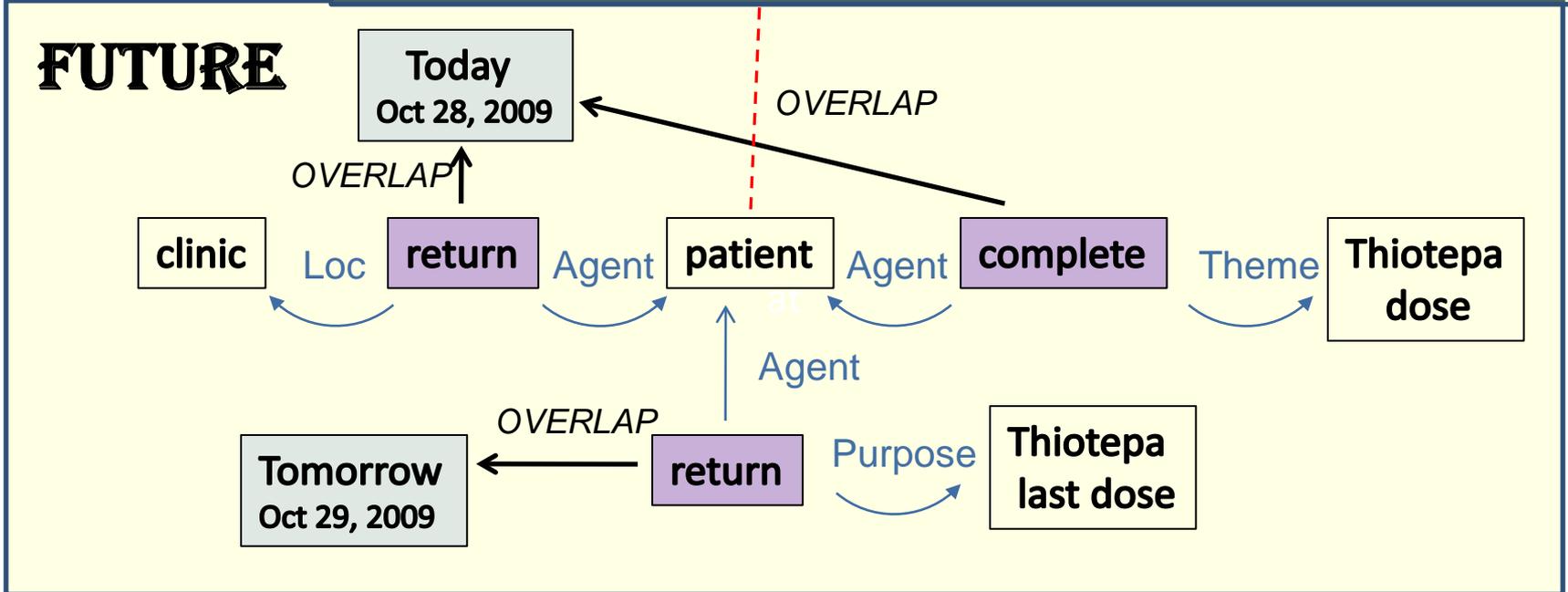
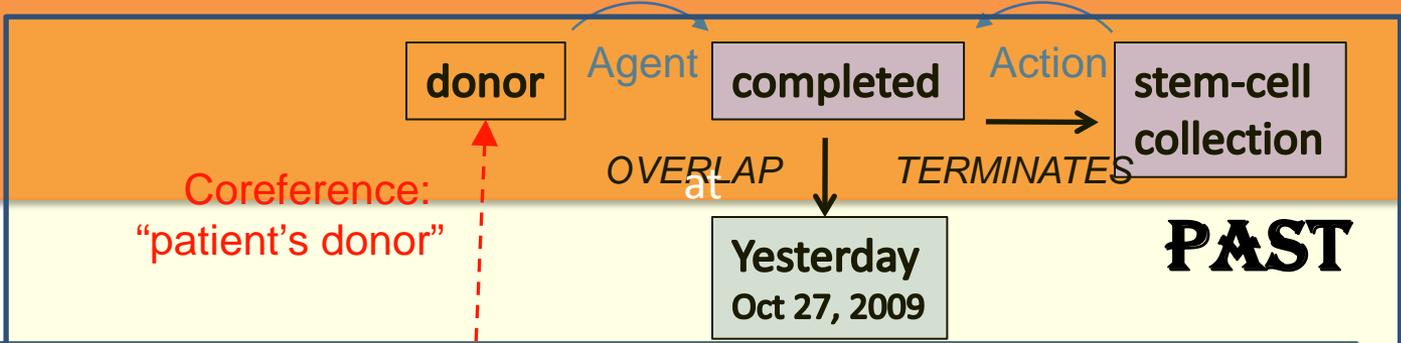


Coreference:
"patient's donor"

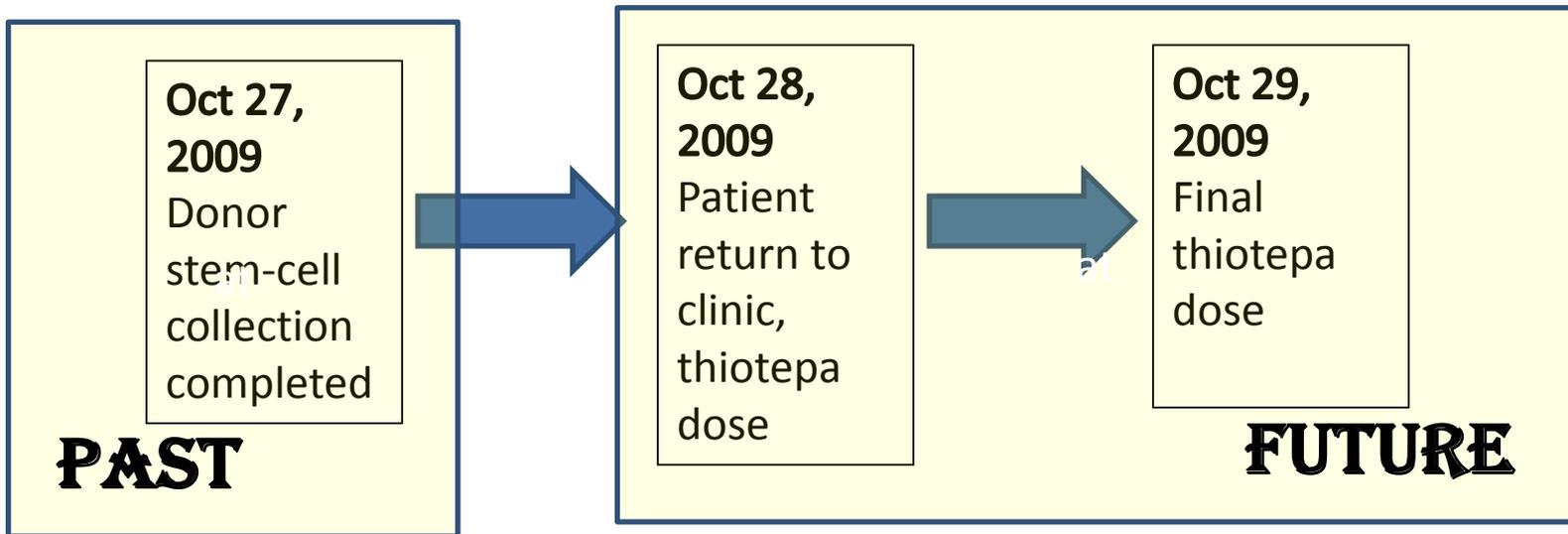


The patient returns to the outpatient clinic today for follow-up
 the patient will complete his thiotepa dose today , and he will return
 tomorrow for the last dose of his thiotepa .

His donor completed stem-cell collection yesterday



The patient returns to the outpatient clinic today for follow-up
 the patient will complete his thiotepa dose today , and he will return
 tomorrow for the last dose of his thiotepa .
 His donor completed stem-cell collection yesterday



The patient returns to the outpatient clinic today for follow-up the patient will complete his thiotepa dose today , and he will return tomorrow for the last dose of his thiotepa .

His donor completed stem-cell collection yesterday



Conclusions

- Linguistic Modeling is a critical component in the development of expressive specification languages (annotation markups)
- The elements of the specification are the foundation for feature identification, for both supervised and semi-supervised machine learning algorithms.
- The problem of “feature discovery” is classical linguistic modeling over BIG DATA.