The Tarsqi Toolkit
Temporal Modeling for the news domain and the medical domain

Marc Verhagen, Brandeis University
Presentation to the AMIA NLP Working Group
Washington, October 2012
Tarsqi People

- James Pustejovsky (Brandeis)
- Inderjeet Mani (Georgetown University, Mitre)
- Roser Saurí, Bob Knippen, Ben Wellner, Jessica Moszkowicz, Alex Plotnick, Amber Stubbs, Russell Entrikin
TimeML
Several kinds of information are need to answer temporally charged questions.

Yesterday Holly was running a marathon when she twisted her ankle. David had pushed her.

1. When did the running occur?  
   Yesterday.
2. When did the twisting occur?  
   Yesterday, during the running.
3. Did the pushing occur before the twisting?  
   Yes.
4. Did Holly keep running after twisting her ankle?  
   Probably not.
TimeML

- **Time expressions**: times, dates, durations, sets
- **Events**: any eventuality, process or state that can be located in time
- **Subordinating links**: reporting verbs, perception verbs, intentional processes and states, modality
- **Temporal links**: temporal relations between events and times
Features of TimeML

• Identifies time expressions;
  – **Time points**: yesterday, at 5pm
  – **Durations**: three years
  – **Sets**: every Saturday

• Identifies signals determining interpretation of temporal expressions;
  – **Temporal Prepositions**: for, during, on, at;
  – **Temporal Connectives**: before, after, while.

• Identifies event expressions;
  – **tensed verbs**: has left, was captured, will resign;
  – **stative adjectives**: sunken, stalled, on board;
  – **event nominals**: merger, Military Operation, Gulf War;

• Creates dependencies between events and times:
  – **Anchoring**: John left on Monday.
  – **Orderings**: The party happened after midnight.
  – **Embedding**: John said Mary left.
Temporal Expressions

Time Points

<TIMEX3 tid="t1" type="TIME" value="T24:00">midnight</TIMEX3>
<TIMEX3 tid="t2" type="DATE" value="2005-02-15" temporalFunction="TRUE" anchorTimeID="t0">tomorrow</TIMEX3>

Durations

<TIMEX3 tid="t6" type="DURATION" value="P2W" beginPoint="t61" endPoint="t62">two weeks</TIMEX3> from <TIMEX3 tid="t61" type="DATE" value="2003-06-07">June 7, 2003</TIMEX3>
<TIMEX3 tid="t62" type="DATE" value="2003-06-21" temporalFunction="true" anchorTimeID="t6"/>

Sets

<TIMEX3 tid="t1" type="SET" value="P1M" quant="EVERY" freq="P3D">three days every month</TIMEX3>
<TIMEX3 tid="t1" type="SET" value="P1M" freq="P2X">twice a month</TIMEX3>
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.
Events

- An event is something that occurs in time
- Some states are not events in TimeML
- Taken in the most general sense and not limited to a limited number of classes (as for example with the ACE classes)
- The argument structure is not part of the event
- Implicit distinction between the event and the interval it occupies
- Generics not included
Event Classes

- **Occurrence:**
  - die, crash, build, merge, sell, take advantage of, ..
- **State:**
  - Be on board, kidnapped, recovering, love, ..
- **Reporting:**
  - Say, report, announce,
- **I-Action:**
  - Attempt, try, promise, offer
- **I-State:**
  - Believe, intend, want, ...
- **Aspectual:**
  - begin, start, finish, stop, continue.
- **Perception:**
  - See, hear, watch, feel.
Different Notions of Events

- **Topic**: “well-defined subject” for searching
  - document- or collection-level
- **Template**: structure with slots for participant named entities
  - document-level
- **Mention**: linguistic expression that expresses an underlying event
  - phrase-level (verb/noun)
MAXICARE HEALTH PLANS INC and UNIVERSAL HEALTH SERVICES INC have dissolved a joint venture which provided health services.

Systems can fill such templates at ~60% accuracy from news texts (MUC evaluations)
Events in TimeML

- **Mention**: linguistic expression that expresses an underlying event
  - Phrase-level (verb/noun)
- Since they correspond to surface mentions, easier to annotate and recognize
  - Accuracy is > 88% (ARDA AQUAINT (TARSQI))
- Like templates
  - they are linked to times
- Unlike templates
  - the times are **resolved**
    - 87% accuracy in time resolution (TERN evals: timex2.mitre.org)
    - the links involve temporal relations
  - the events are temporally ordered
  - the links also involve other logical relations (subordinating and aspectual)
SLINKS

Subordination Links are used for contexts introducing relations between two events, of the following sort:

**Factive:** Certain verbs introduce an entailment (or presupposition) of the argument's veracity. They include *forget* in the tensed complement, *regret, manage*:

- *John forgot that he was in Boston last year.*
- *Mary regrets that she didn’t marry John.*

**Counterfactive:** The event introduces a presupposition about the non-veracity of its argument: *forget (to), unable to* (in past tense), *prevent, cancel, avoid, decline*, etc.

- *John forgot to buy some wine.*
- *John prevented the divorce.*

**Evidential:** Evidential relations are introduced by REPORTING or PERCEPTION:

- *John said he bought some wine.*
- *Mary saw John carrying only beer.*

**Negative evidential:** Introduced by REPORTING (and PERCEPTION?) events conveying negative polarity:

- *John denied he bought only beer.*

**Conditional:** Introduced when a conditional is present:

- *If Mary leaves today, John will leave tomorrow.*
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.

Note the chain of slinks at the bottom: said-allowed-leave
TLinks

• Goal is to order events
• Temporal relations between
  – Two events
  – An event and a time (anchoring)
  – Two times (automatic)
• As informative as possible in the service of question answering
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.

Linking a timex to the document creation time
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.

Linking an event to the document creation time
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.

Linking an event to the nearest timex
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.

Linking two events in a subordination relation
Iraq's Saddam Hussein, *facing* U.S. and Arab troops at the Saudi border, *today* sought *peace* on another front by *promising* to *withdraw* from Iranian territory and *release* soldiers *captured* during the Iran-Iraq *war*. Also today, King Hussein of Jordan *arrived* in Washington *seeking* to *mediate* the Persian Gulf *crisis*. President Bush on *Tuesday* *said* the United States may *extend* its naval *quarantine* to Jordan's Red Sea port of Aqaba to *shut off* Iraq's last *unhindered* trade route.

In another *mediation effort*, the Soviet Union *said* *today* it had *sent* an envoy to the Middle East on a series of *stops* to *include* Baghdad. Soviet officials also *said* Soviet women, children and invalids would be *allowed* to *leave* Iraq.

*Linking two main events of consecutive sentences*
Iraq's Saddam Hussein, facing U.S. and Arab troops at the Saudi border, today sought peace on another front by promising to withdraw from Iranian territory and release soldiers captured during the Iran-Iraq war. Also today, King Hussein of Jordan arrived in Washington seeking to mediate the Persian Gulf crisis. President Bush on Tuesday said the United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba to shut off Iraq's last unhindered trade route.

In another mediation effort, the Soviet Union said today it had sent an envoy to the Middle East on a series of stops to include Baghdad. Soviet officials also said Soviet women, children and invalids would be allowed to leave Iraq.
Financing Business:

Blockbuster Entertainment Corp.

10/27/99 WALL STREET JOURNAL (3) BY LOSS MARKET NEWS (SON) BLOCKBUSTER ENTERTAINMENT CORP.

said it raised $92 million from an offering of liquid yield option notes. The gross proceeds from the sale of the notes, which will be due on Nov. 1, 2004, will be used to reduce existing debt and for general corporate purposes, the company said.

The debt reduction is expected to save the Fort Lauderdale, Fla., home video concern about $2 million a year in interest expense. The zero-coupon subordinated notes have no periodic interest payments. Each note is being offered at $308.32 per $1,000 principal amount at maturity, representing an 8% yield to maturity. In addition, each note can be converted into Blockbuster Entertainment common stock at a rate of 13.851 shares per note.

Merrill Lynch Capital Markets Inc.
Tarsqi Toolkit (TTK)
Tarsqi Toolkit Architecture

text document

DocumentModel

PreProcessing

Entities

Evita

BTime

Events

Times

Temporal Relations

SLinks

Slinket

S2T

Blinker

Classifier

Link Merging

TLinks

SputLink
GUTime

- Tags TIMEX3 expressions
- Built on TempEx, licensed from MITRE
  - TempEx handles an early version of TIMEX2
- Extends TempEx
  - Handles durations
  - Includes begin and end points
    - Includes option to distinguish speech time from DCT
  - Handles variety of MODs
  - Handles basic TIMEX3 temporal functions
  - Handles Euro date formats
- [http://complingone.georgetown.edu/~linguist/GU_TIME_DOWNLOAD.HTML](http://complingone.georgetown.edu/~linguist/GU_TIME_DOWNLOAD.HTML)
BTime

- Problems with GUTime/Tempex
  - Perl script using complex regular expressions
  - Performed reasonably well, but hard to maintain, here is one of the simpler regular expressions used by GUTime

```perl
if($string =~ /quarter/io) {
    $string =~ s/($OT+($TENumOrds|$TEOrdinalWords)$CT+\s+ $OT+quarter$CT+)/
               <TIMEX$tever TYPE="DATE">$1</TIMEX$tever>/sogi;
    ... }
```
Btime - Grammar

• Context-free grammar implemented using the Early algorithm
• Grammar with 82 declarative rules

```
date ->
    day month year { CalendarDate(_[2],_[1],_[0]) } |
    month day year { CalendarDate(_[2],_[0],_[1]) } |
    day month "," year { CalendarDate(_[3],_[1],_[0]) }
```
Btime – Temporal Functions

• take an underspecified timex3 and create the normalized value given an anchor time, which is determined using a set of simple heuristics.

str(today) : 2011-10-13
day_of_week(today) : THU
PastAnchoredTimePoint(WeeksDuration(3))(today) : 2011-09-22
FutureAnchoredTimePoint(WeeksDuration(3))(today) : 2011-11-03
Increment(Day)(today) : 2011-10-14
Increment(Week)(today) : 2011-W42-1
Decrement(Day)(today) : 2011-10-12
Decrement(Week)(today) : 2011-W40-1
NextInstance(DayOfWeek(5))(today) : 2011-10-20
Evita

- Event Recognition System
- Locates and tags all event-referring expressions in the input text that can be temporally ordered
- Identifies grammatical features associated with the event-denoting expressions which inform about temporal and modal information
  - Tense, aspect, polarity, modality, event class
- Event Durations (Jerry Hobbs and Feng Pan)
Slinket

- Partial modal parser which identifies subordination constructions that introduce modality information in text
- Specifically, infinitival and ‘that’ clauses embedded by:
  - Factive Predicates (‘regret’)
  - Perception Predicates (‘see’)
  - Events of attempting (‘try’), volition (‘want’), command (‘order’), …
Beyond Slinket – De Facto

• Roser Sauri
• Identifying factuality sources

(1) **Nixon claimed** that White House counsel **John Dean** had conducted an investigation into the Watergate matter and **found** that **no-one from the White House was involved**.

(2) **What members of the White House were involved in the Watergate matter?**

(3) **No-one**
Temporal Links

• **S2T**
  – Take an SLINK and derive a TLINK
  – “He said he will be teaching on Monday”
  – Reporting SLINK + past tense + future progressive $\rightarrow$ Before TLINK

• **Blinker**
  – Rule based, six rule sets
  – Time-time, event-DCT, time-event, event-event

• **The Classifier**
  – MaxEnt machine learning algorithm
TLINK Merging

- Temporal links can (and will) be inconsistent
  - $X < Y$ & $Y < Z$ & $Z < X$

- Temporal links are very sparse
  - Manual annotation
    - Anywhere from 80-99% of possible temporal links will be missed
  - Automatic annotation
    - Slightly higher coverage but lower precision

- Need component to ensure consistency and fill in the gaps
Temporal Closure (1)
Temporal Closure (2)
Temporal Closure (3)
Merging Procedure

• Allen’s constraint propagation
  – Events are placed in a graph and the arcs are labeled by disjunctions over the basic relations
  – New constraints on an arc are propagated through the graph using a 13x13 transitivity table that models the transitive behavior of all pairs
  – Similar to path consistency and graph closure algorithms

• Add TLINKs from previous components one by one and check consistency at each step
Visualization
Displaying Event Anchoring

- Temporal information by color-coding events
- Problem: if two events have the same color, then you assume they are at the same time
- Use a bar with a spectrum that spans the whole TimeML graph in a document

Distinguish three bins: past, present and future
Assign parts of the spectrum to these bins
Displaying Event Anchoring

• Four kinds of events with simple scheme
  
  past       present           future
  unknown

  1987       DCT

• Scales up nicely when more precise information is available
Pax can’t find buyers. She estimates her properties, worth a hundred thirty million dollars in October, are worth only half that now. They believe ah it will be always up going up and up ah forever. Nobody believe thin any more. Of all of Asia’s economies, Hong Kong is the most robust. But in the past three months, stocks have plunged, interest rates have soared and the downturn all across Asia means that people are not spending here. Hotels are only thirty percent full. You can get seventy percent discounts at the shopping malls. Three thousand dollar pearls for eight hundred dollars. A two hundred dollar wool jacket for fifty dollars. Still, there are few buyers.
T-Box

Unlike Tango, there are strict rules on where to place an event that:

1. precedes another event (left-to-right arrows)
2. includes an another events (box inclusions)
3. Is simultaneous with another event (box stacking)
TBox Example
TBox Example

constructio...

believe(e26)

worth(e25)
  * October(t8)

going(e27)

Now(t7)
  now(t9)

lucky(e19)

backed(e18)

five_year(t3)
  hit(e6)

fallen(e4)
  the_last_tw...
  lost(e5)
Tempeval
Evaluating Temporal Relations

- Dependencies between relations
- Syntactic vs semantically

- Temporal closure
TempEval Solution

- Split the task into manageable parts
- Take away the dependencies in the annotation
- Independent tasks where all annotation instances inside each task are independent from the other instances in the task
- For each task, we can use pair wise comparisons
Tempeval2 Tasks

C. Determine the temporal relation between an event and a time expression in the same sentence. For TempEval-2, this task is further restricted by requiring that either the event syntactically dominates the time expression or the event and time expression occur in the same noun phrase.

D. Determine the temporal relation between an event and the document creation time.

E. Determine the temporal relation between two main events in consecutive sentences.

F. Determine the temporal relation between two events where one event syntactically dominates the other event. This refers to examples like "she heard an explosion" and "he said they postponed the meeting". However, we are investigating whether for some tasks the more precise set of TimeML relations could be used.

before, after, overlap, before-or-overlap, overlap-or-after, vague
72% IAA for anchorings, 65% for orderings
Combining relations from different task and running temporal closure creates a connected graph.
<table>
<thead>
<tr>
<th>team</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIPSem</td>
<td>0.81</td>
<td>0.59</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TIPSem-B</td>
<td>0.81</td>
<td>0.59</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>JU_CSE</td>
<td>0.63</td>
<td>0.80</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>NCSU-indi</td>
<td>0.63</td>
<td>0.68</td>
<td>0.48</td>
<td>0.66</td>
</tr>
<tr>
<td>NCSU-joint</td>
<td>0.62</td>
<td>0.21</td>
<td>0.51</td>
<td>0.25</td>
</tr>
<tr>
<td>TIPSem</td>
<td>0.55</td>
<td>0.82</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>TIPSem-B</td>
<td>0.54</td>
<td>0.81</td>
<td>0.55</td>
<td>0.60</td>
</tr>
<tr>
<td>TRIOS</td>
<td>0.65</td>
<td>0.79</td>
<td>0.56</td>
<td>0.60</td>
</tr>
<tr>
<td>TRIPS</td>
<td>0.63</td>
<td>0.76</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>USFD2</td>
<td>0.63</td>
<td>-</td>
<td>0.45</td>
<td>-</td>
</tr>
</tbody>
</table>
Chronicles, Trajectories and Entities
Event-based Chronicles

- Persons
- Organizations
- Fray Bartolome Center
- Peace and Justice
- PRI
- Red Masks
- Roman Catholic Church
- Zapatistas

### 1994
- uprising
- declared
- launched
- war
- rebellion

### 1996
- early 1996
- February
- signed

### Peace talks
- break_down
- stalled

### Recent months
- *12/22/97
- accused
- mediate

Zapatistas declared war...

...Roman Catholic Church mediates between government and Zapatistas

Zapatistas signed peace accord...

...break down of peace talks between Zapatistas and government...

Rebels accused state officials...

Zapatistas launched rebellion...
### Chronicle Example

**Places:**
- Chiapas
- Tuxtla Guiterrez
- San Cristobal de Las Casas
- Ocosingo
- **Agua Azul**
- Ruins of Palenque
- Border with Guatemala

**People:**
- David
- Gregg

<table>
<thead>
<tr>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike(David, to_A.A.)</td>
<td>David biked 30 miles to the town of Agua Azul.</td>
</tr>
<tr>
<td>Arrive(Gregg, A.A.)</td>
<td>David was surprised Gregg had not arrived.</td>
</tr>
<tr>
<td>Bike(Gregg, to_A.A.)</td>
<td>David biked alone towards Agua Azul. A few miles from Agua Azul, Gregg was</td>
</tr>
<tr>
<td>Robbed(Gregg, @A.A.)</td>
<td>suddenly robbed while climbing a steep hill.</td>
</tr>
<tr>
<td>Surprise(David, Gregg)</td>
<td>(Doc1): bike(David, to_A.A.) BEFORE surprise(David, Gregg)</td>
</tr>
<tr>
<td></td>
<td>(Doc1): surprise(David, Gregg) AFTER arrive(Gregg, A.A.)</td>
</tr>
<tr>
<td></td>
<td>(Doc2): bike(Gregg, to_A.A.) INCLUDES climb(Gregg)</td>
</tr>
<tr>
<td></td>
<td>(Doc2): bike(A.A., Gregg) INCLUDES robbed(Gregg, @A.A.)</td>
</tr>
<tr>
<td></td>
<td>(Doc1+Doc2): robbed(Gregg, @A.A.) BEFORE surprise(David, Gregg)</td>
</tr>
</tbody>
</table>
Narrative Containers
Narrative Times

• Narrative Time is the current temporal focus of the document, it is updated while the reader progresses through the document.

• Narrative Container is an interval that extends into the past from the creation time of a document and that can be used to anchor recent events to.
April 25, 2010 7:04 p.m. EDT

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.
Electronic Health Records
Neonatology:
Patient is now 36 h old term infant sent to NICU for eval of tachypnea because of episode of tachypnea with feeds in NN.

Patient is 3.875 kg product of term gestation born by c-section for MRHT on Friday afternoon. Prenatal screens unremarkable.
CRB*: Maternal temp to 100.8 during labor. Did well in OR [**Last Name (Title) 43**] [**Name (Title) 44**] 7, 9.

IN NN has been asymptomatic. Breastfeeding well. Episode of tachypnea to 80 in this morning. Sent to NICU for [**Female First Name (un) **].

On exam pink active non-dysmorphic infant. Well perfused, 100% 
stuated in RA. Skin w/o lesions. Lungs clear. ECG 64152 w/o
arr rhythm. Abdomen benign. Neuro non-focal and age appropriate.

A. Well appearing term infant with one peripartum sepsis risk factor and episode of tachypnea.
P CBC diff BC.
Monitor in NICU til CXR returns.
If cbc normal and patient remains asx here will return to NN for further care and monitoring.

NPN Septic evaluation:
Infant was brought from NBN for evaluation of RR and septic work up. [**Name8 (MD) 29**] MD note for details. Tags checked with
NBN and report given. Infant placed on warmer and exam done and normal. Infant infant stable in RA, RR, HR and BP stable.
Examined by MD [**First Name (Title) 18**] [**Last Name (Title) 42**]. CBC and Blid cx sent. CS 62. CBC normal, bld cx pending.
Ordered to send to NBN. Will check tags and give report.
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**.

12:00 am  ¬food & ¬drink  →  surgery  →  food & drink  →  recovery
HOSPITAL COURSE: After her admission, the patient was evaluated and she progressed with rematch workup.

On November 15 1999, the patient was randomized to the Surgical arm of study.

On November 15 1999, the patient underwent placement of a vented electric LVAD.

Post-op course was significant for bleeding and thrombocytopenia requiring packed red blood cells and platelets; ATN requiring CBVHD with continued need for hemodialysis at the present time via a Groshong catheter; atrial fibrillation requiring cardioversion on November 27 1999; respiratory decompensation following extubation requiring a tracheostomy on December 4 1999 and eventual closure of the tracheostomy on January 4 2000.

Patient was also evaluated for fluctuating mental status and was diagnosed with toxic metabolic encephalopathy with resolution.
Medical Events

• Persistent states
  – TimeML events must be temporally interesting (that is, states must be related to a temporal expression or identifiably change during the course of the document)
  – Therefore, events such as *diabetes* are not captured

• Medical terms
  – Medical terms such as *thrombocytopenia* and abbreviations such as *ATN* are not in the dictionary and will not be recognized.

• Fuzzy times
  – Temporal expressions in the medical domain are generally fuzzy: 18 months could really mean around 18 months, give or take about 6 months.
Now what?

- Temporal Processing for Medical Discharge Summaries
- Subbs & Harshfield 2010
- Two-year NIH project
- Annotating 50-100 discharge summaries
- Adapting event recognition and temporal linking
- Build cross-document temporal database of medical events
Annotation

• Data
  – De-identified free text EHRs from the MIMIC II database, available through PhysioNet.org (Neamatullah, 2008; Goldberger et al, 2000).
  – REALIST study

• Update TimeML and CLEF guidelines

• Clinical E-Science Framework (Sheffield)
  – Entities: condition, intervention, investigation, result, drug or device, locus
  – Relations: has_target, has_finding, has_indication, has_location
EHR Events

• For nominal disambiguation, Evita uses libraries based on Wordnet, Semcor and Timebank.

• UMLS SPECIALIST Lexicon and the UMLS Metathesaurus.
  – using both normalization and typing

• Additional (small) libraries created after annotation
EHR Relations

• The rule sets of Blinker do not include all of the lexical and syntactic constructions that are used liberally in the medically domain. The rule sets was expanded to deal with specific constructions like status post MI 1983.

• The classifier will be trained on the manually annotated data.
Example record

HPA: This 87y old female with PMH of COPD has a [**12-31**] day history of increasing SOB and decreasing mental status. She came to [**Hospital1 2**] EW with family and initially had say of 97% on 4L NP. Her sats though decreased along with mental status and she was intubated. X-ray showed LLL pneumonia and she had WC of 18. BP dropped after intubation and she received 1 liter NS and another liter was hung. She received 500mg levoquin and solumedrol 125mg iv. She received multiple doses of IV ativan for sedation. She was transfered to MICU intubated and sedated for further care.

ALLERGIES: Note chart says PCN, but on further investigation that is not so. So NKDA

PMH: COPD on chronic home 02. Rectal adenocarcinoma, S/P low ant resection in 3/06. Anxiety. SIADH. Migraines. Osteoporosis.
Some of the major issues confronted when integrating multiple patient narratives into a chronicle of medical events from the different narratives include the following:

- entity co-reference (that is, reference to the same patient);
- references to the same event from different records;
- discovering the temporal orderings between events from different records;
Patient Chronicle

• CSP for identity of events
• To discover the temporal orderings between events from different records, we need to do the following:
  – Create a document containing all records for a particular patient;
  – Generate temporal links between all temporal expressions that denote a time or date;
  – Apply temporal closure to the document.
Patient Chronicle

Record 1
- March 9th
- e1_condition → e2_intervention

Record 2
- March 11th
- e3_condition

Records 1 and 2
- March 9th
- e1_condition → e3_condition
- e2_intervention
- March 11th
Initial experiment - Statins

• Amber Stubbs & BJ Harshfield
• Given 300+ EHRs for people selected to participate in the R3\(^i\) realist medical study (one record per person)
• Find out whether a majority of the people were on statins or other lipid-lowering drugs at the time they were admitted to the hospital.
• Data were not de-identified
  – Software development done on i2b2 data
  – Later run on the actual R3\(^i\) data to ensure HIPAA compliance.
Proposed solution

- Using scripts to identify medication names and the Tarsqi Toolkit to identify relevant times and create links.
- Find links between statins and admission times, using the relation type to answer the question as to whether the medication was being taken before admission, or not.
Headers in EHRs

- Non-standard sections such as *Medical History, Course of Treatment, and Discharge Medications*, with multiple variations on each possible name
  - Typing headers and ordering them temporally makes it easier to temporally locate events relative to the admission date.
- Use a script that finds section headers in documents and associates them with Timex3 tags with type="HEAD" and a mod of past, present, after, or not (for lists of allergies).
- Modified Blinker to keep track of the current header value, so links between events and the admission date could be labeled.
- Used a hack to ensure that pre-processor does the right thing with headers (improved accuracy by nearly 10% overall).
Medication names

• Evita does not have the vocabulary to identify most medical events or medication names
  – medications probably shouldn't be considered events anyway, but calling them events was the simplest solution to the problem.

• Since we were only looking for a small number of medication names, we developed a post-Evita Perl script to find each mention of the medications we were looking for and label them as events with type=MEDICATION.
Actual procedure

• Identify section headers and the admission date
• Run TTK, using only the preprocesser, GUTime and Evita.
• Identify the statins and label them as events
• Identify headers and associate them with timexes
• Run TTK, now only using a specialized version of Blinker
• Extract all temporal anchorings and orderings for the statin events
• Manual evaluation
Results

• The program split the texts into two categories: "before", for those who were taking statins at the time of admission, and "not" for those who were not taking them at all, taking them only during/after the hospital stay, and for those that were ambiguous.

• The human annotators had the option to label an EHR as "before", "not", or "ambiguous", for those cases that needed more information.

• Accuracy: 88%.
Conclusion

• Modify the TTK to deal with the EHR format:
  – identify lists versus narrative text
  – identify section headers and predict their relation to the admit and discharge dates

• Add new tag types to deal with medications, dosages and frequencies.
  – Medications are now labeled as events, dosages are not labeled, frequencies are occasionally labeled as Timexes.

• New vocabulary lists for medical events, as well as medications, etc.
Thank you