$\{\texttt{tue}, \texttt{apr}, 12_{\texttt{day}}, 2005_{\texttt{year}}\}$

Understanding Times - A Constraint-based Approach

Benjamin Han (benhdj@cs.cmu.edu)

Donna Gates, Lori Levin,
Michael Kohlhase, Alon Lavie, and Jaime Carbonell

Language Technologies Institute
Carnegie Mellon University

The Question

- Would it be nice if we can automatically...
  - Have an email turned into a meeting schedule showing up in my calendar application;
  - Obtain the answer to the question such as "*was there any shuttle mission in '87?*"
  - Summarize events in Beethoven's life, in chronological order

The Question

- But what kind of information do we have?
  - From email: “*I’m free **next week**. Let’s meet **on Monday***”.
  - From newswire: ”*after the Columbia accident in **‘86**, shuttle missions were suspended **in the following 2 years***”
  - From web: “***At 14** Beethoven was able to deputize for his teacher. **Three years later**... Prince Maxmilian Franz sent him to Vienna to further his education...*”
  - From news, intelligence sources, historical data etc.

The Solutions

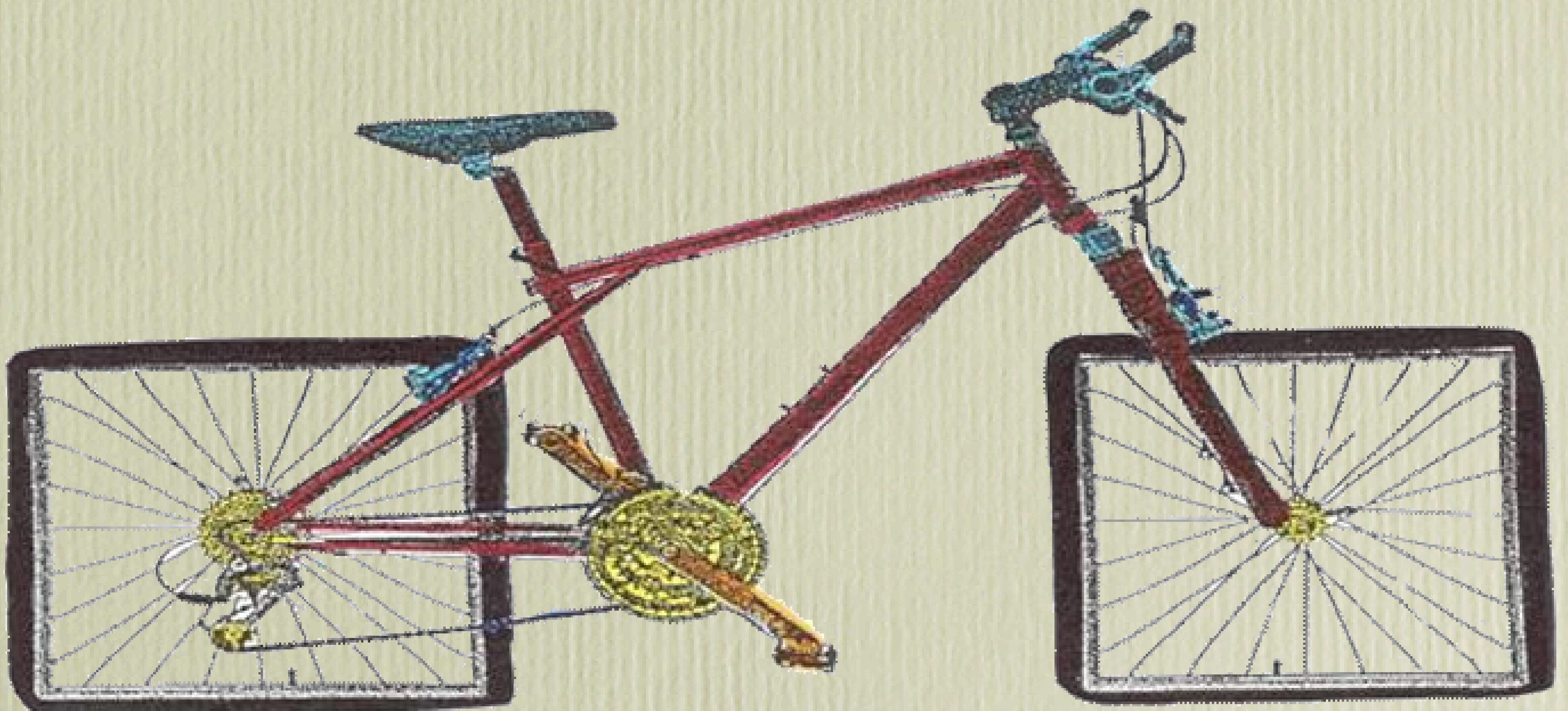
- **Identify**, **represent**, and **reason with** temporal information in NL
  - Verb tense/aspect
  - **Temporal expressions**: "*Wednesday*", "*in a week*", "*current*", "*recently*", "*..., when the market stabilized*", etc.
  - Events

What are we doing here?

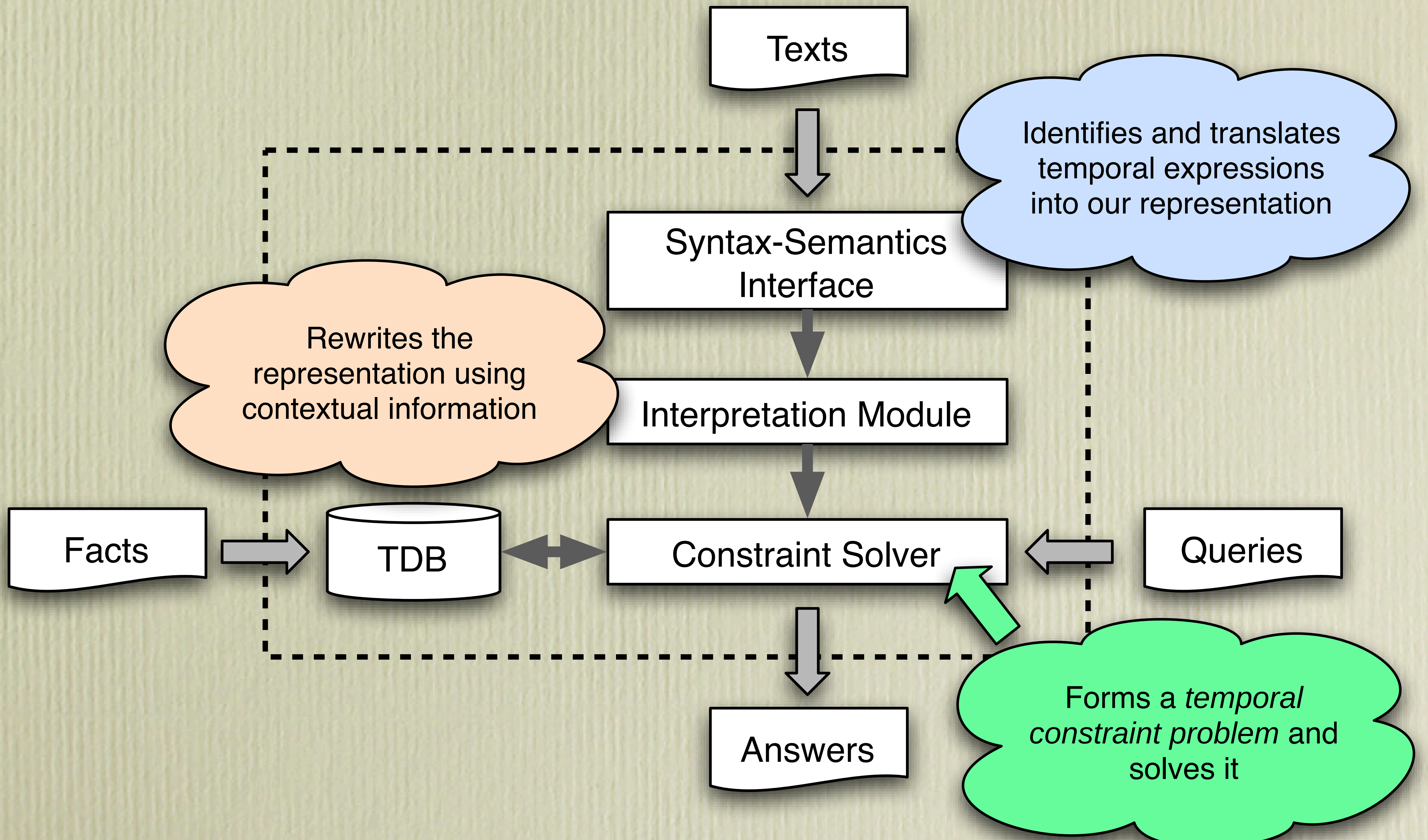
"I wish I could run that fast too!"

Dawn of Human Civilization

Well, not quite...



Our Attack Plan

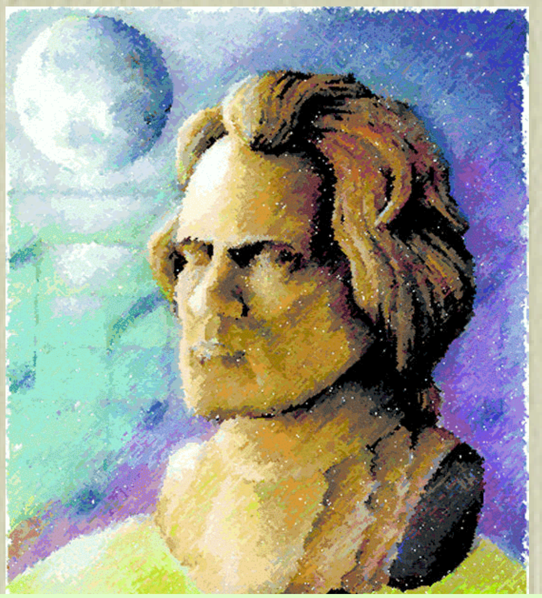


Outline

- Modeling temporal scenarios in temporal constraint satisfaction problems
- Modeling human calendars
- Representing temporal information (Time Calculus)
- Applications

Beethoven's trip to Vienna

T1 : $\{T1' + |14_{year}|\}$



“**At 14** Beethoven was able to deputize for his teacher. **Three years later**, recognizi

T2 : $\{\_ + |3_{year}|\}$

talent, Prince Maximilian Franz sent him to Vienna to further his education. He would soon return **within two weeks** on the news that his mother was dying. She passed away **3 months later on July 17, 1787**.

T3 : $\{\_ + |(< 2)_{week}|\}$

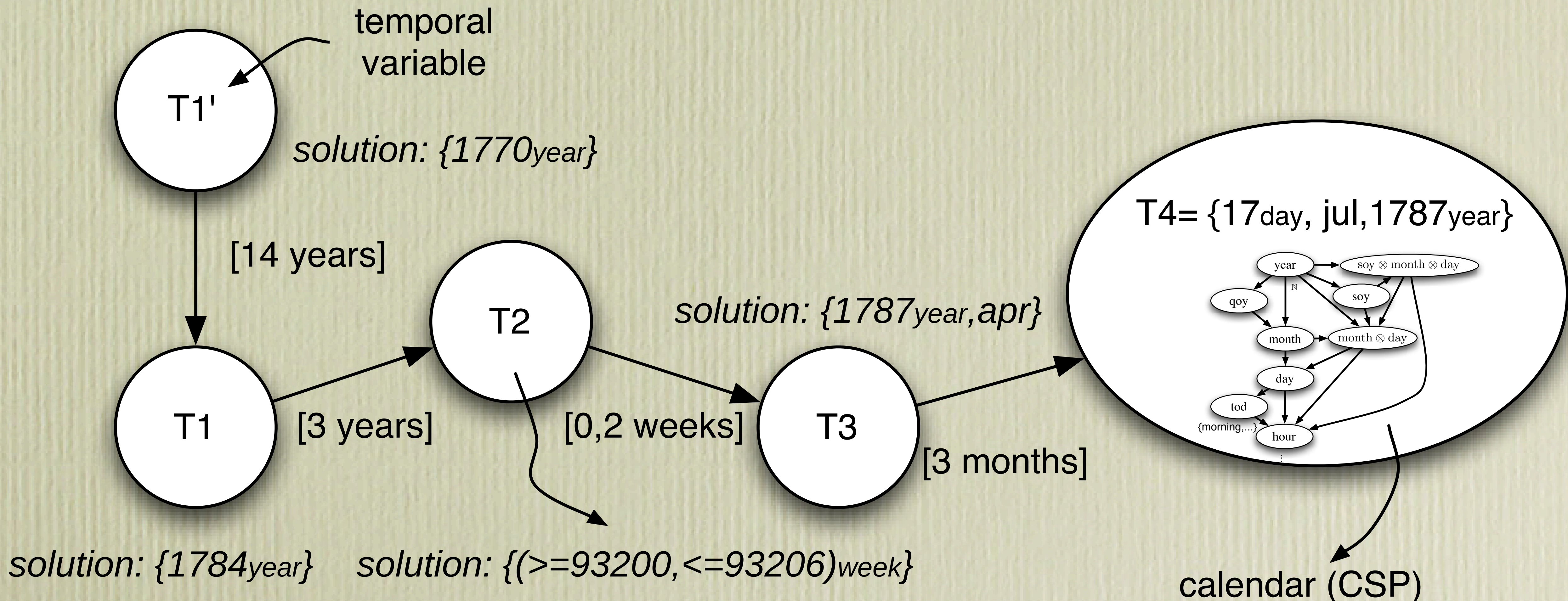
T4 : $\{\_ + |3_{month}|, jul, 17_{day}, 1787_{year}\}$

Instantiating Temporal Foci

$\texttt{T1}: \quad \{\texttt{T1}'+|\texttt{14}_{\texttt{year}}|\}$	*deputizing-at-14*
$\texttt{T2}: \quad \{\texttt{T1}+|\texttt{3}_{\texttt{year}}|\}$	*off-to-Vienna*
$\texttt{T3}: \quad \{\texttt{T2}+|(< 2)_{\texttt{week}}|\}$	*Return*
$\texttt{T4}: \quad \{\texttt{T3}+|\texttt{3}_{\texttt{month}}|, \texttt{jul}, \texttt{17}_{\texttt{day}}, \texttt{1787}_{\texttt{year}}\}$	*Mother's-death*

Temporal Constraint Satisfaction Problems (TCSP)

"Constraint Networks"



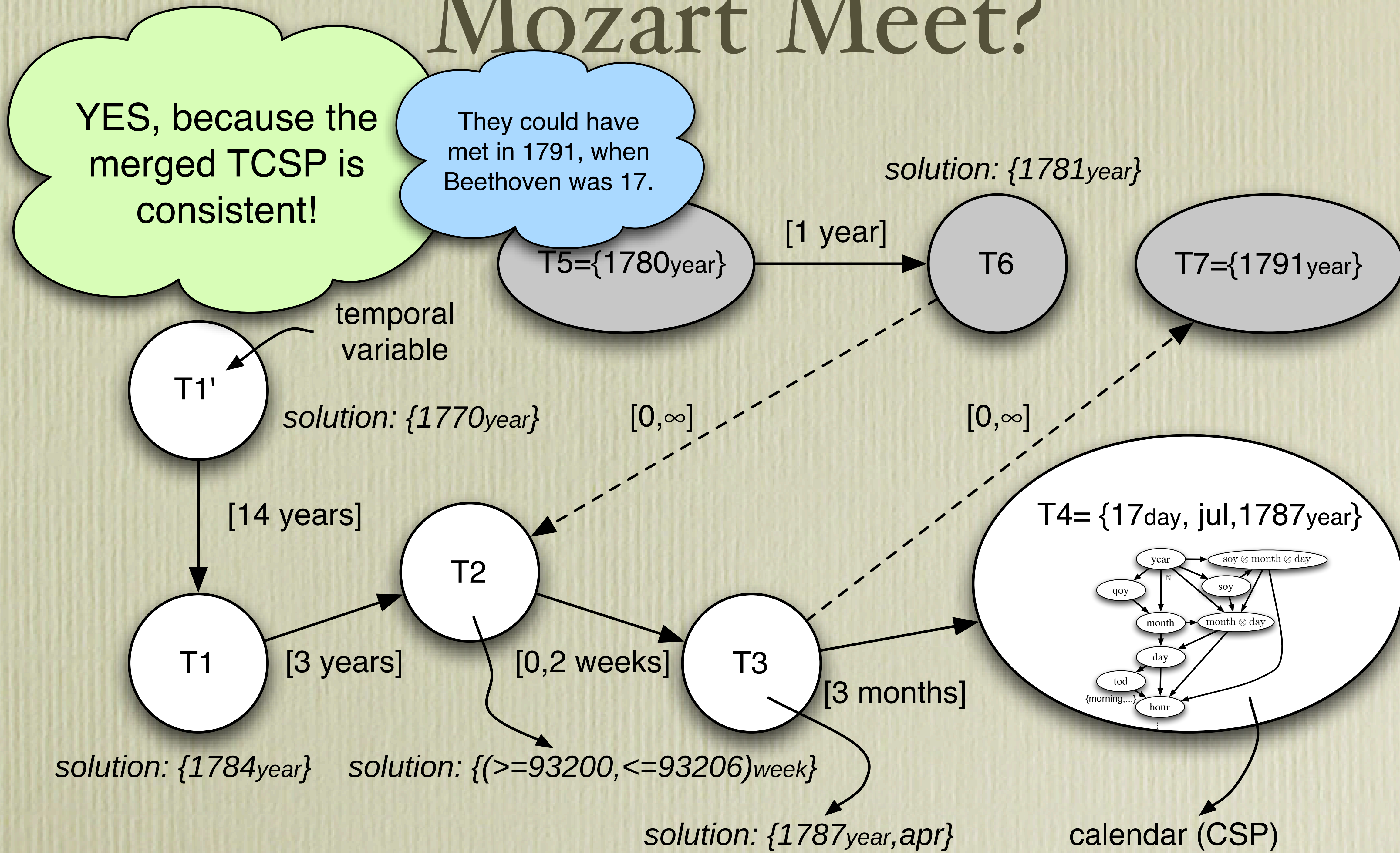
Example: Mozart's Activity in Vienna

"Mozart went to Munich to compose the opera **late in 1780**. **The next year**, he was $T5 : \{1780_{year}\}$

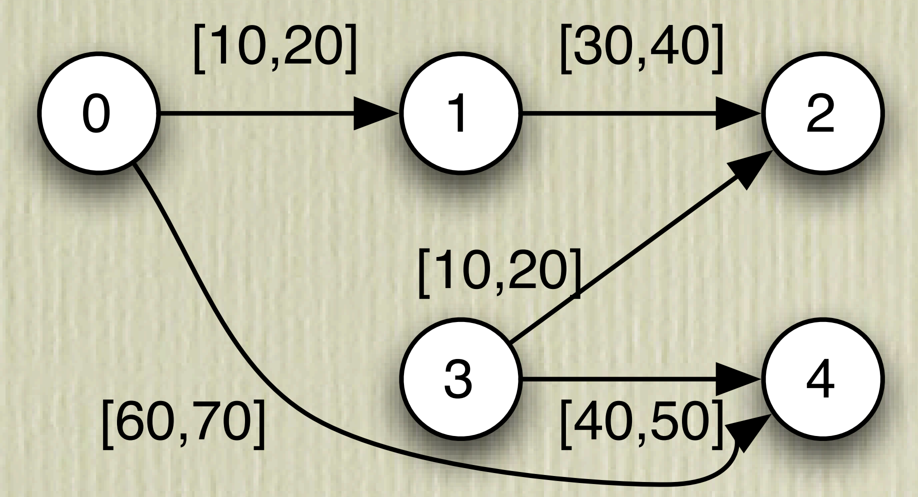
$T6 : \{\_+|1_{year}|\}$ from Munich to Vienna, where the Salzburg court was in residence on the accession of a new emperor. Mozart lived in Vienna for the rest of his life, until he died **in 1791**."

$T7 : \{1791_{year}\}$

Question: Did Beethoven & Mozart Meet?

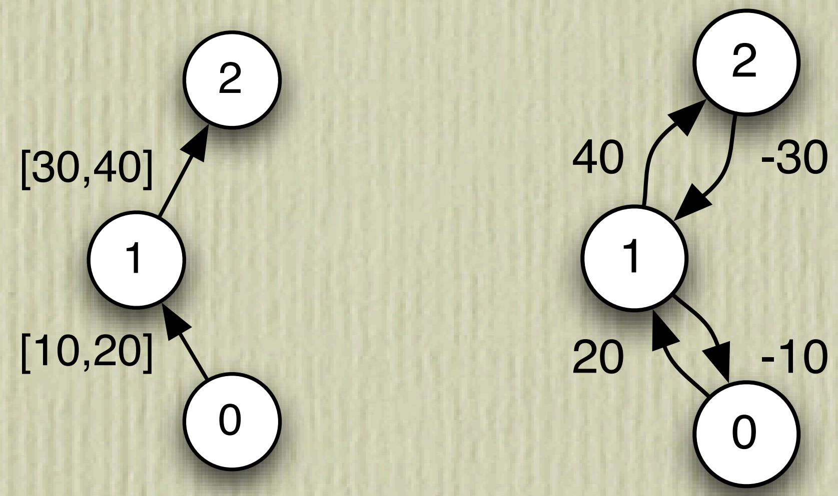


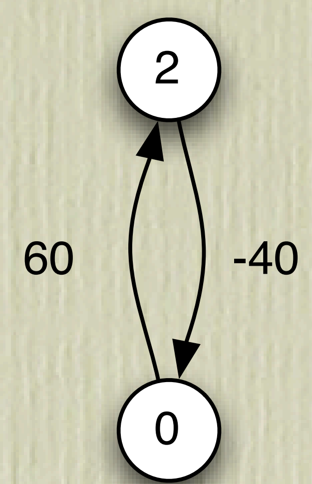
Solving TCSPs {Dechter et al, 91}



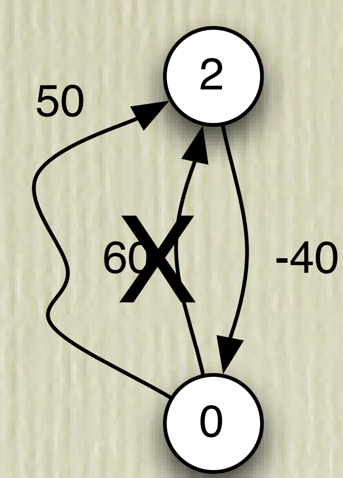
- We get to know...
  - Is it *consistent*?
  - What are the feasible assignments to the variables (nodes)?
  - What are the relations that can possibly hold between two variables?

Solving Simple TCSPs

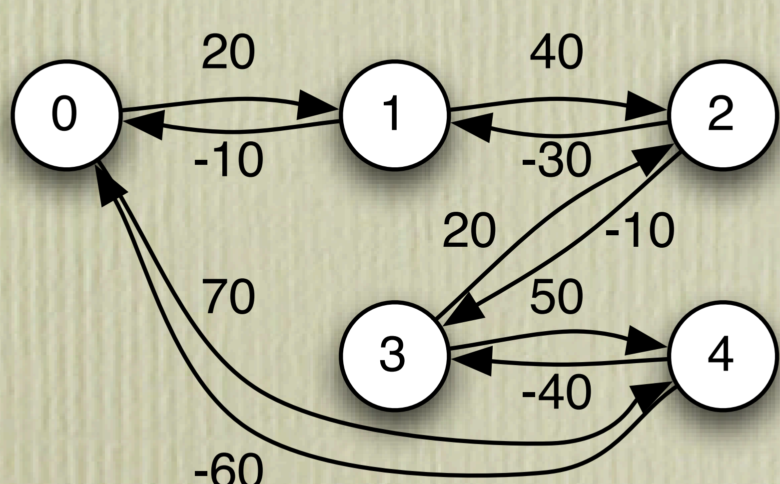




distance graph



minimal network



all-pairs-shortest algorithm

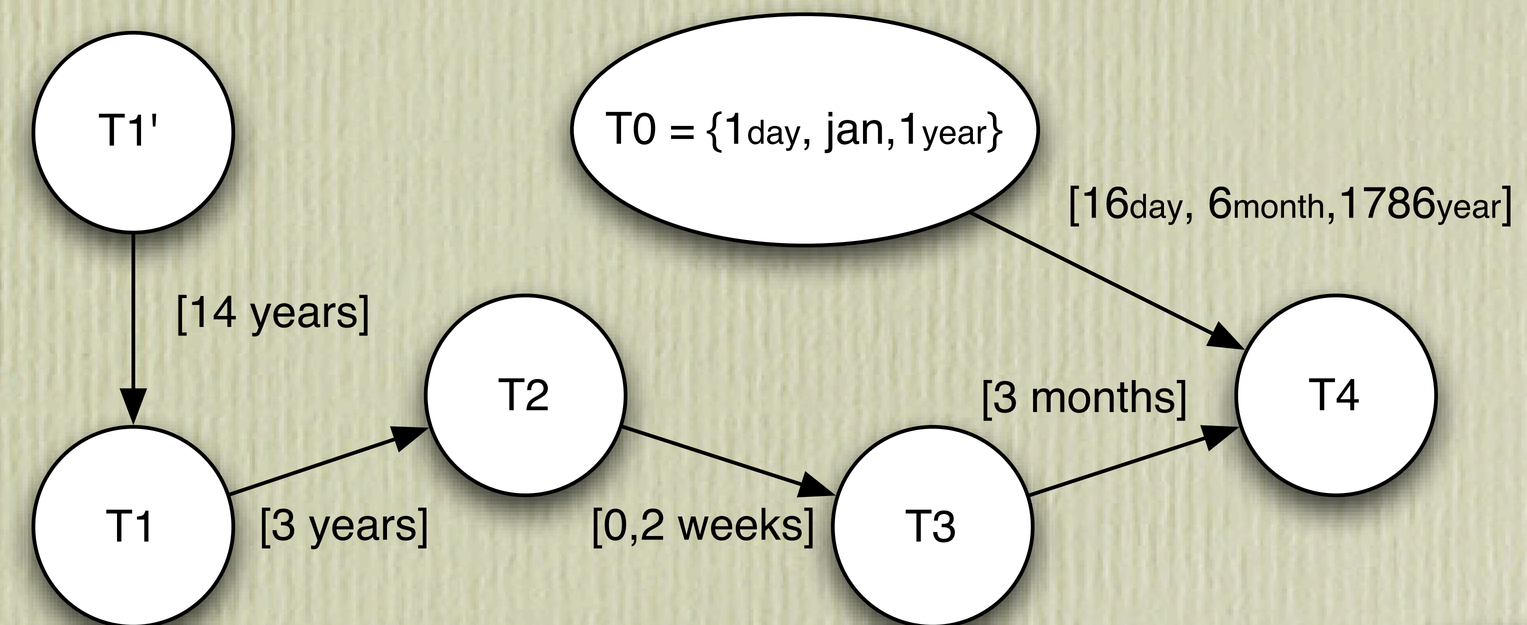
	0	1	2	3	4
0	[0]	[10,20]	[40,50]	[20,30]	[60,70]
1	[-20,-10]	[0]	[30,40]	[10,20]	[50,60]
2	[-50,-40]	[-40,-30]	[0]	[-20,-10]	[20,30]
3	[-30,-20]	[-20,-10]	[10,20]	[0]	[40,50]
4	[-70,-60]	[-60,-50]	[-30,-20]	[-50,-40]	[0]

runs in *polynomial* time ($O(n^3)$) !

Solving General TCSPs

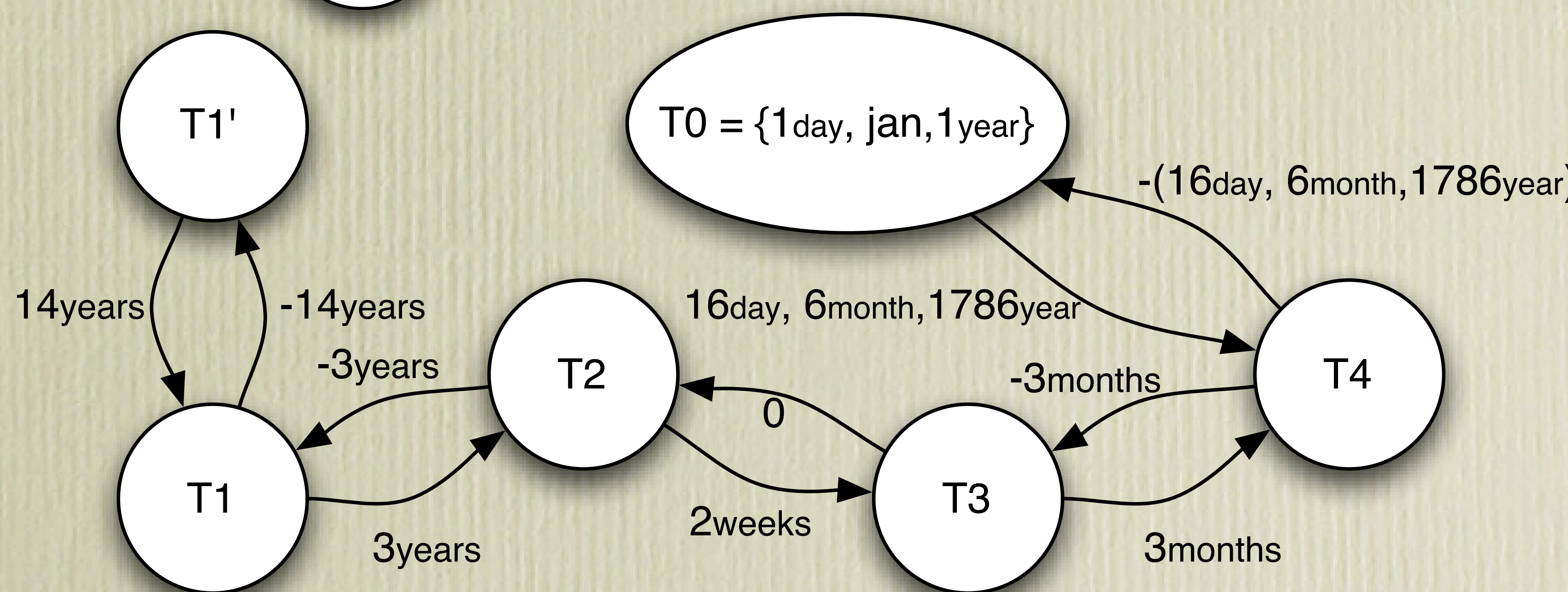
- Can have disjunctive constraints (labels on the edges)
- It's an NP-hard problem (even with max. 2 labels per edge) $O(n^3 k^e)$
- In practice we can use path-consistency and certain heuristics to improve efficiency.

Back to Beethoven vs. Mozart



Constraint network

distance graph



And the Answers are...

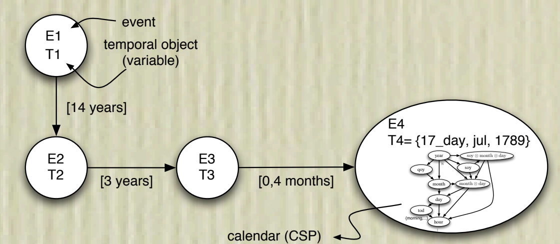
d = 16 days, 6 months and 1786 years

$$
\begin{aligned}
T_4 - T_0 &: \quad [d] \\
T_3 - T_0 &: \quad [d - 3_{month}] \\
T_2 - T_0 &: \quad [d - 3_{month} - 2_{week}, d - 3_{month}] \\
T_1 - T_0 &: \quad [d - 3_{month} - 2_{week} - 3_{year}, d - 3_{month} - 3_{year}] \\
T_1' - T_0 &: \quad [d - 3_{month} - 2_{week} - 3_{year} - 14_{year}, d - 3_{month} - 3_{year} - 14_{year}]
\end{aligned}
$$

$$
\begin{aligned}
T_4 &= \text{July 17, 1787} \\
T_3 &= \text{April 1787} \\
T_2 &= \text{between 93200th and 93206th week (or between March 18 and May 5)} \\
T_1 &= 1784 \\
T_1' &= 1770
\end{aligned}
$$

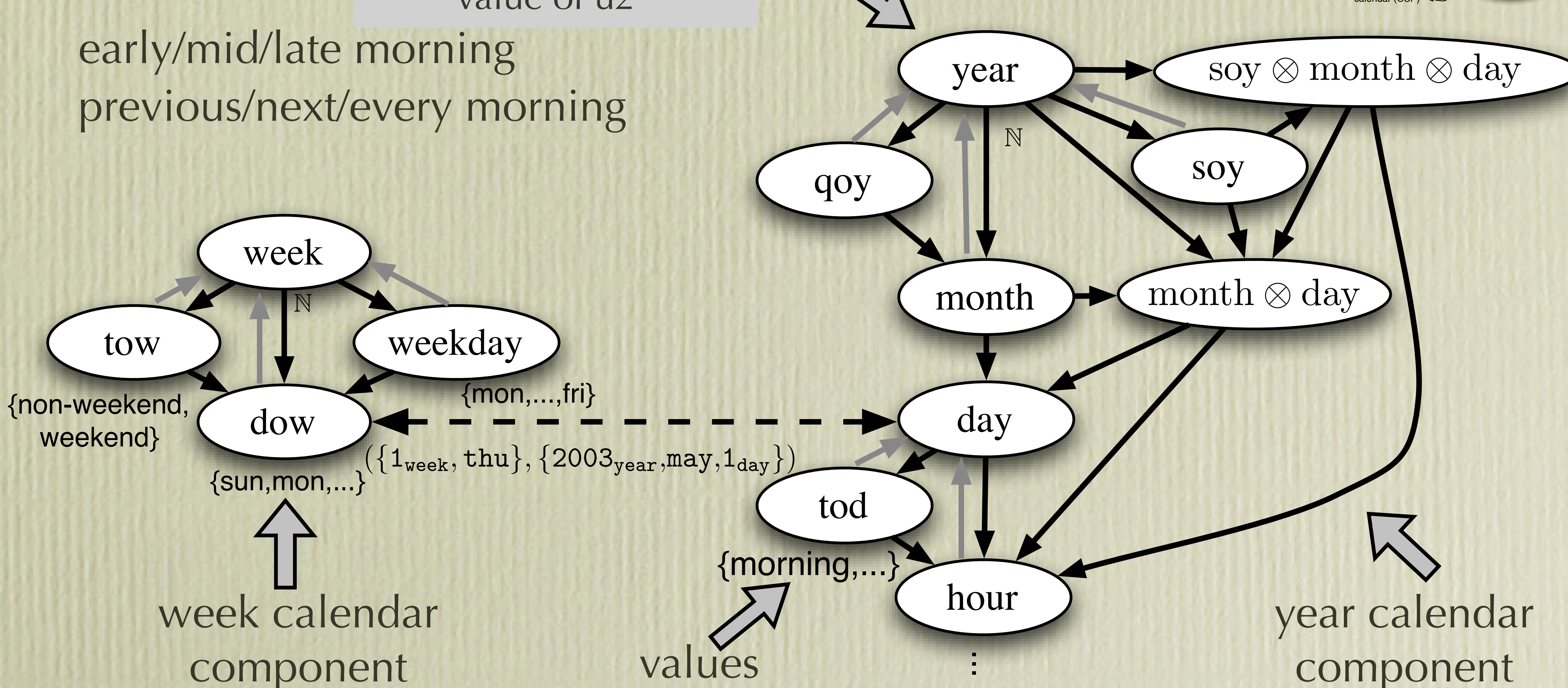
Wait, how do you know the answer to "*x* days - *y* weeks"? "*x* months - *y* fiscal years"?

Modeling Human Calendars as Constraint Systems



u1 is periodic in u2 iff every value of u1 is in the cover of every value of u2

early/mid/late morning
previous/next/every morning



Cover mappings: $C_{\texttt{qoy},\texttt{month}}(1_{\texttt{qoy}}) = \{\texttt{jan},\texttt{feb},\texttt{mar}\}$

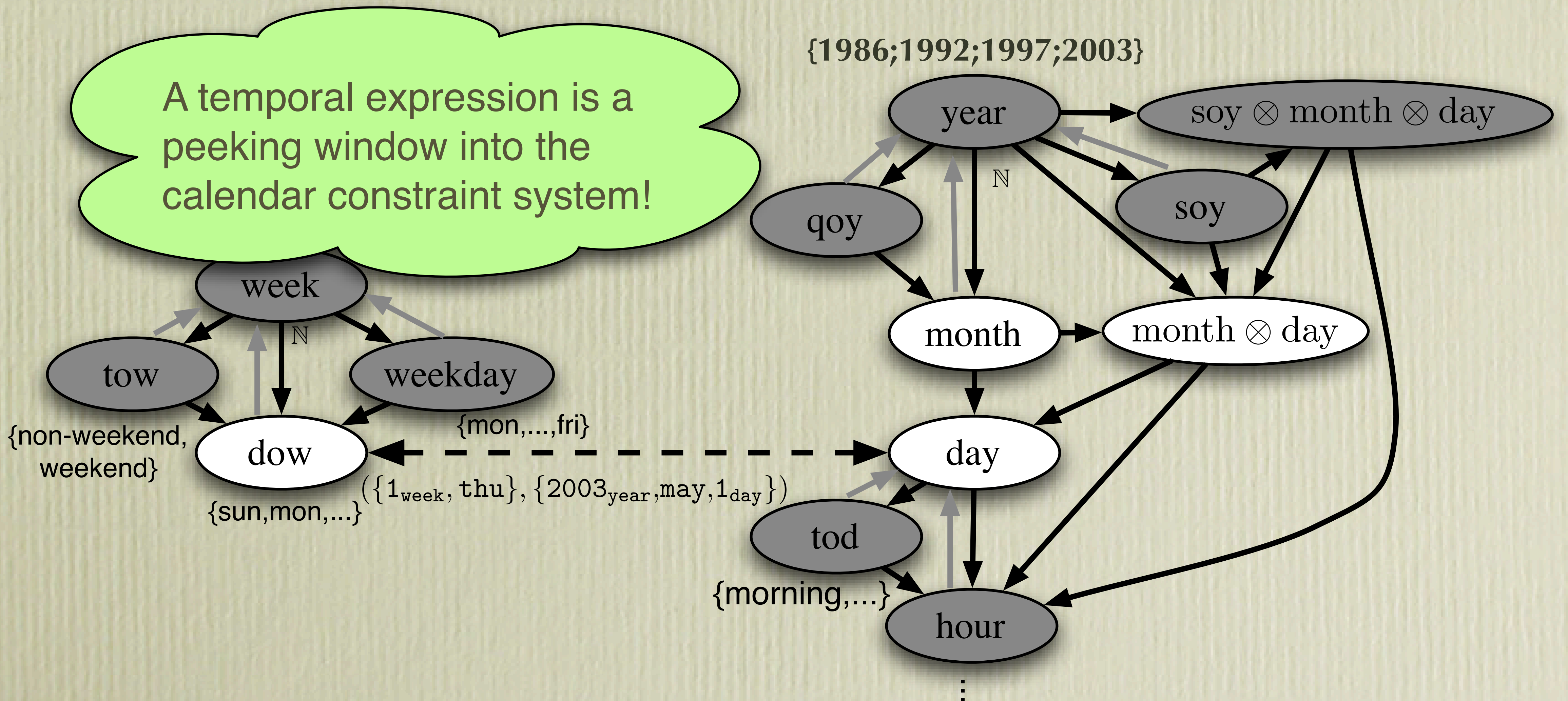
Time Detective

"Over the long ***holiday weekend honoring Martin Luther King Jr.'s birthday****, the mission managers su*

the 3rd Monday in January

and some participants took time off. They met on ***the next day****,* ***Jan. 21****, five days into the flight."*

"The first shuttle mission flew in ***1981****."*

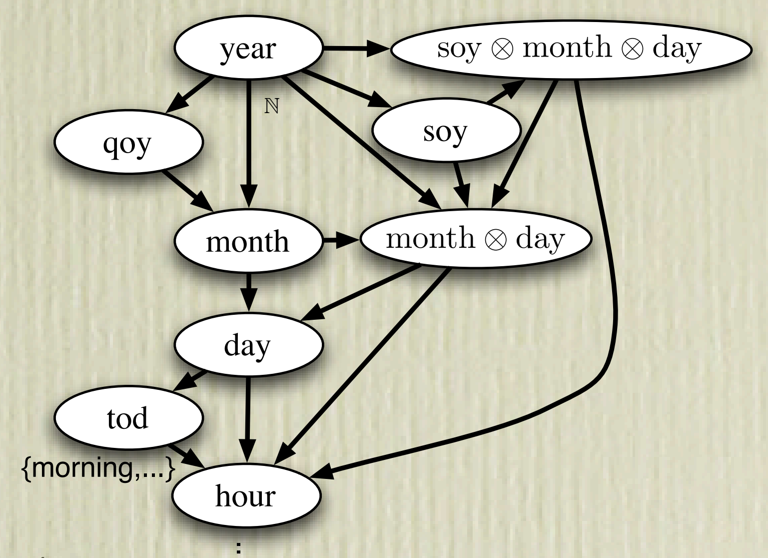


Other Things that Calendars Do

- Which of the following times is the earliest?

$c_1 = \{\texttt{1995}_{\texttt{year}}, \texttt{1}_{\texttt{qoy}}\}$ $c_2 = \{\texttt{1995}_{\texttt{year}}, \texttt{apr}\}$ $c_3 = \{\texttt{1995}_{\texttt{year}}, \texttt{feb}\}$

$c_3 < c_2$ $c_1 < c_2$ $c_1 \; ? \; c_3$



Rule #1: only coordinates based on the same grounded chain are comparable.

Rule #2: *x*<*y* to be true if and only if every possible *x* is earlier than every possible *y*

- Add 2 seconds to 23:59:59 to Feb 28, 2004?
  - Generate and test
  - Speed-up is possible via periodicity relations and pre-computing

How Complex can Temporal Expressions be?

- "***Sept. 9, 1987***": Anchorable **coordinate**, but may come in different granularity.
- "***Wednesday***": Doubly under-specified: (i) which Wednesday? (ii) point or interval?
- "***4 o'clock***": Ambiguity.
- "***today***": Referring to a reference time.

More Expressions...

- "***last week***": Shifting with a certain amount of time with <u>granularity change</u>.
- "***last Wednesday***": Amount to shift is not expressed in temporal units.
- "***an hour and 15 minutes***": A **quantity**.
- "***the 3rd week of November***": Ordinal expression.

And More...

- "***Tuesday and Thursday***": An **enumeration**, not logical conjunction.
- "***from now until 1995***": A specialized enumeration I - an interval expression.
- "***every week in May***": A specialized enumeration II - a recurrence expression.
- "***twice a week***": A rate expression.

And More...

- Pragmatic expressions
  - "***about*** ***midday*** ***New York time***"
  - "***late*** ***this year***"
  - "***most*** ***of the day***"
  - "***one*** ***recent*** ***day***"
- Non-standard units
  - "***this*** ***term***"
  - "***every*** ***weekday***"
  - "***the previous*** ***fiscal year***"

And Even More!

- Things are even nastier!
  - "***at least 30 days***": Inequalities.
  - "***4 1/2 hours a day***": Rate expressions + quantities.
  - "***every six days***": Recurrence + quantities.
  - "***at least 2 hours every second Monday from June to August, 1987***".
  - ...

Introducing Time Calculus

- A typed language with 3 types of "objects":
  - **Coordinate** (*C*): Points at certain granularity;

    $\{\texttt{1987}_{\texttt{year}}, \texttt{sep}, \texttt{9}_{\texttt{day}}\}$

  - **Quantity** (*Q*): Polarity neutral duration;

    $|\texttt{1}_{\texttt{hour}}, \texttt{30}_{\texttt{min}}|$

  - **Enumeration** (*E*): Set of points (including intervals);

    $[\{\texttt{tue}\}, \{\texttt{thu}\}]$

Connectives and Negation

- For conjunctions: ‘,’

$\{1987_{\text{year}}, \texttt{sep}, 9_{\text{day}}\}$

- For disjunction: ‘;’

$\{1987_{\text{year}}; 1988_{\text{year}}\}$

- For ambiguity: ‘|’

$\{4_{\text{hour}} | 16_{\text{hour}}\}$

- For negation: ‘!’

$\{1987_{\text{year}}, \texttt{sep}, !9_{\text{day}}\}$

Operators and Relations

- Using *operators* to construct complex objects
  - *fuzzy* shifting: $\{2003_{\texttt{year}}, \texttt{feb}, 28_{\texttt{day}}, 8_{\texttt{hour}}\} + |1_{\texttt{day}}|$
  - intervals: $[\{\texttt{may}\} : + |2_{\texttt{month}}|]$
  - ordinals: $\{|2_{\texttt{sun}}| @ \{\texttt{may}\}\}$
  - *pattern* recurrence: $[\{*\texttt{wed}, \texttt{may}\}]$
  - *arithmetic* recurrence: $[\{\texttt{bi}\ 1896_{\texttt{year}}\} / |4_{\texttt{year}}|]$
- Using relations to construct complex objects

$\{(>= 4, <= 5)_{\texttt{hour}}\}$

More on Enumerations

- "*9am to 5pm, but not 2pm*"

$$[[\{9_{\texttt{hour}}\}:\{17_{\texttt{hour}}\}]\backslash[\{14_{\texttt{hour}}\}]]$$

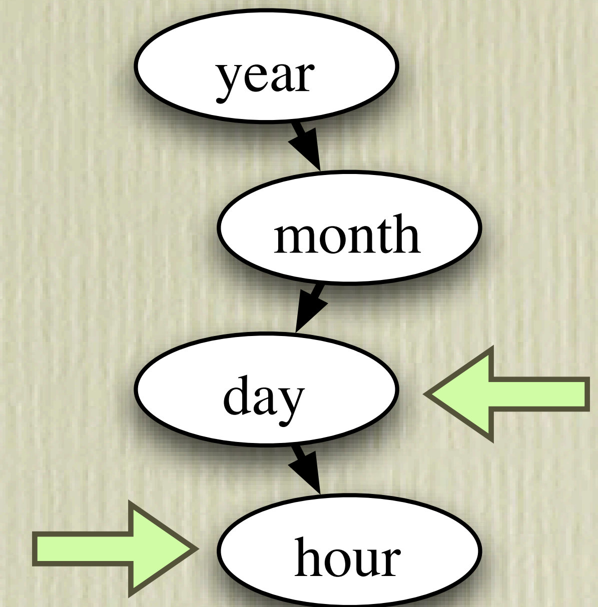
- "*the 5th Wednesday in every month*"

$$[|5_{\texttt{wed}}|@\{*_{\texttt{month}}\}]$$

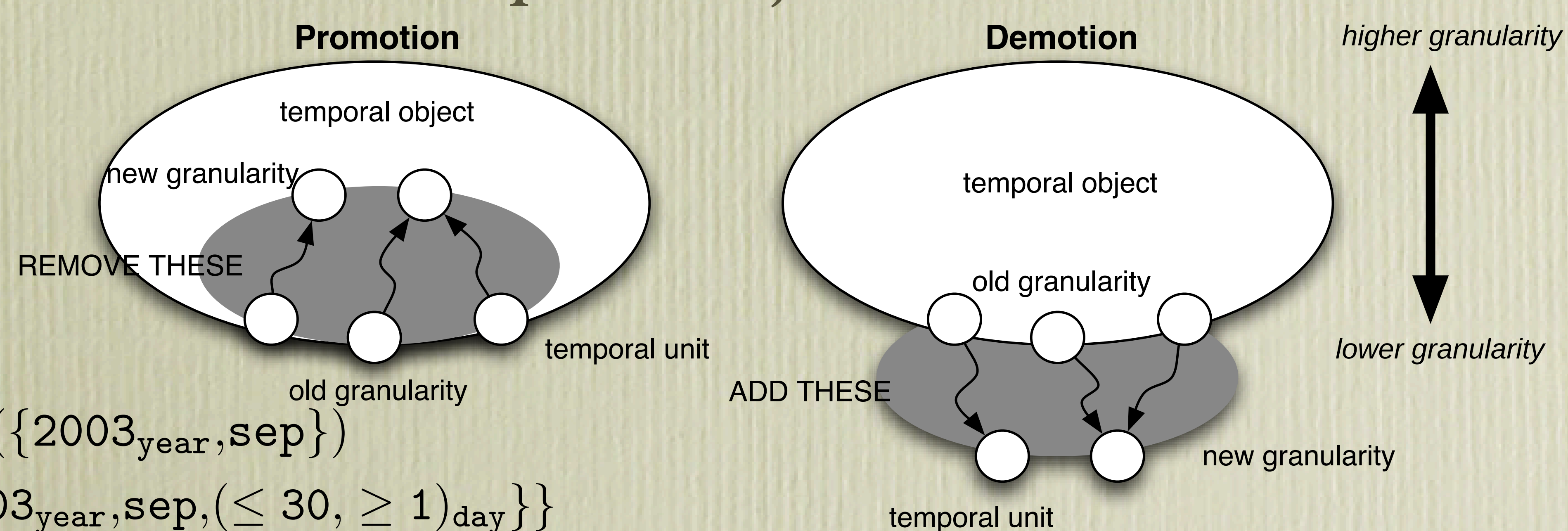
Granularity

- Why does it matter?

$$\{2003_{\mathtt{year}}, \mathtt{feb}, 28_{\mathtt{day}}, 8_{\mathtt{hour}}\} + |1_{\mathtt{day}}|$$



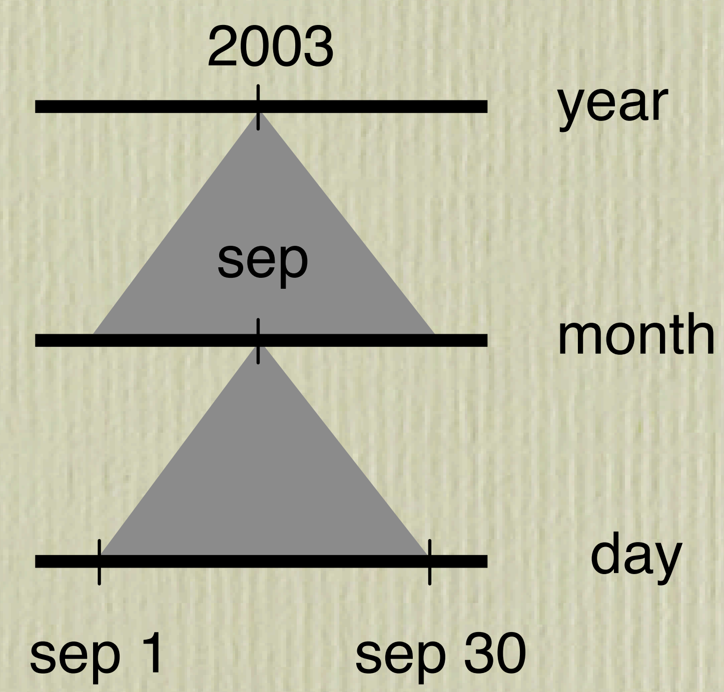
- What is it?
  - A set of minimal units.
- Granularity change: installing a new set of minimal units into a temporal object

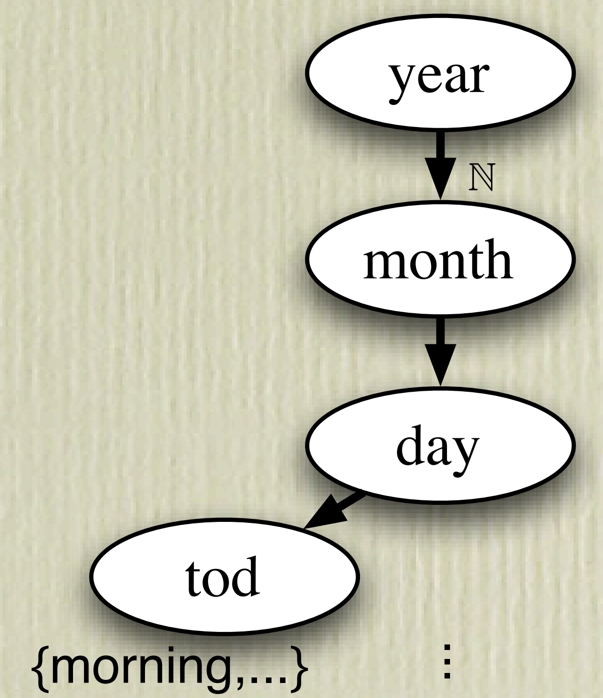


$$\rightarrow_{\{\mathtt{day}\}} (\{2003_{\mathtt{year}},\mathtt{sep}\})$$
$$= \{\{2003_{\mathtt{year}},\mathtt{sep},(\leq 30, \geq 1)_{\mathtt{day}}\}\}$$

Re-interpretation

- Re-interpreting a coordinate into an enumeration





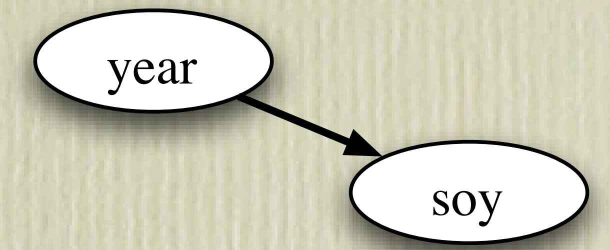
$$C \rightarrow E_{\{\mathtt{day}\}}(\{\mathtt{2003}_{\mathtt{year}},\mathtt{sep}\}) = [\mathtt{i}[\min(\rightarrow_{\{\mathtt{day}\}}(\{\mathtt{2003}_{\mathtt{year}},\mathtt{sep}\})) : \max(\rightarrow_{\{\mathtt{day}\}}(\{\mathtt{2003}_{\mathtt{year}},\mathtt{sep}\}))]]$$

$$= [\mathtt{i}[\{\mathtt{2003}_{\mathtt{year}},\mathtt{sep},\mathtt{1}_{\mathtt{day}}\} : \{\mathtt{2003}_{\mathtt{year}},\mathtt{sep},\mathtt{30}_{\mathtt{day}}\}]]$$

$$C \rightarrow E_{\{\mathtt{year}\}}(\{\mathtt{2003}_{\mathtt{year}},\mathtt{sep}\}) = [\{\mathtt{2003}_{\mathtt{year}}\}]$$

Typed Operators

- Granularity is rolled into types
  - type of $\{\texttt{2005}_{\texttt{year}}, \texttt{summer}\}$ is $C_{\{\texttt{soy}\}}$



- Operators are typed
- Type coercion kicks in behind the scene for granularity change and re-interpretation.

	Type	Meaning	Example
$+/-$	$\mathcal{E}_{\rightarrow_{\mathrm{g(u}(op_2))}} \times \mathcal{Q}_{\mathrm{g}(op_2)} \rightarrow \mathcal{C}_{\mathrm{g}(op_2)}$	forward/backward fuzzy shifting	$\{\_+\|1_{\texttt{month}}\|\}$ ("*next month*")
$++/--$	$\mathcal{E}_{\rightarrow g_m} \times \mathcal{Q}_{\rightarrow g_m} \rightarrow \mathcal{C}_{g_m}$ $g_m = \min_{\mathcal{M}}(\mathrm{g}(op_1) \cup \mathrm{g}(op_2))$	forward/backward exact shifting	$\{\_++\|1_{\texttt{month}}\|\}$ ("*exactly one month after*")

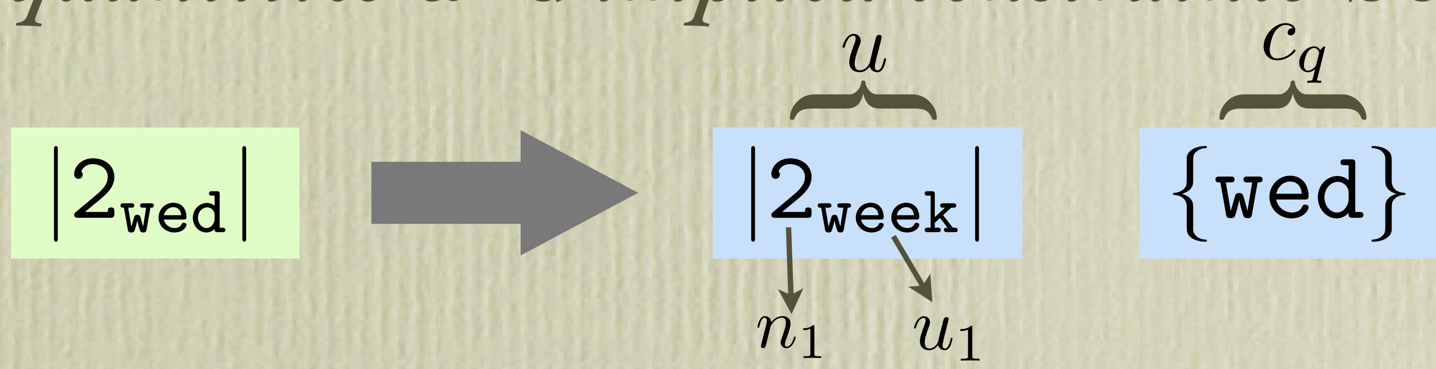
⋮

Example: Fuzzy Shifting

- Type of Fuzzy Shifting operator

  $E_{\rightarrow_{g(\mathrm{u}(op_2))}} \times Q_{g(op_2)} \rightarrow C_{\rightarrow_{g(op_2)}}$

- Quantities are broken down into *pure-unit quantities* and *implied constraints* before shifting



- Definition of the operator

  $e + q \;:=\; \{\max(e[-1]) \oplus_{u_1} n^1 \ldots \oplus_{u_m} n^m, c_q\}$

Example: Fuzzy Shifting

$$e + q \ := \ \{\max(e[-1]) \oplus_{u_1} n^1 \ldots \oplus_{u_m} n^m, c_q\}$$

$$\begin{aligned}
_+|2_{\texttt{wed}}| &= \{2005_{\texttt{year}}, \texttt{summer}\}+|2_{\texttt{wed}}| \\
&= \{\max(\mathcal{C} \rightarrow \mathcal{E}_{\{\texttt{week}\}}(\{2005_{\texttt{year}}, \texttt{summer}\})[-1]) \oplus_{\texttt{week}} 2, \texttt{wed}\} \\
&= \{\max([\mathbf{i}[105487_{\texttt{week}}{:}104600_{\texttt{week}}]][-1]) \oplus_{\texttt{week}} 2, \texttt{wed}\} \\
&= \{\{104600_{\texttt{week}}\} \oplus_{\texttt{week}} 2, \texttt{wed}\} \\
&= \{104602_{\texttt{week}}, \texttt{wed}\} \\
&= \{2005_{\texttt{year}}, \texttt{sep}, 28_{\texttt{day}}\}
\end{aligned}$$

Application I: Project RADAR

- Building agents that read your emails
- Meeting-scheduling emails are converted into "templates"
- Agents negotiate meeting times using the templates

Annotating Times in Emails

"*Classes begin **Monday 25**.*" (sent at: 01:17 Tue, May 6, 1997)

$\{\texttt{mon}_{\texttt{dow}}, 25_{\texttt{day}}\}$ | 1997–05–25 ?:?:? – 0500

"*Please put it in the directory **by 5pm wednesday.***"
(sent at: 11:41 Tue, Aug 26, 1997)

$\{(\mathbf{b\ i})\{\{|1_{\texttt{wed}}|@\{\mathbf{bi}\ \_\}\}, 17_{\texttt{hour}}\}\}$

[?–?–? ?:?:? – 0500 : 1997–08–27 17:00:00 – 0500]

"*I'm scheduled to teach on **Wednesday evening from 6:00 to 9:00.***" (sent at: 01:32 Tue, Sep 2, 1997)

$[\{|1_{\texttt{wed}}|@\{\mathbf{bi}\ \_\}, 18_{\texttt{hour}}\}:\{21_{\texttt{hour}}\}]$

[1997–09–03 18:00:00 – 0500 : 1997–09–03 21:00:00 – 0500]

Building Templates

- Input: emails
- Output: meeting templates

"I'd like to meet with you ***sometime next week****. I'm free* ***on Tuesday and Thursday****."* (sent at: 21:43 May 5, 2004)

```
:
(timestamp"2004-05-05[21:43-5]")
:
(time-slots
  (time-slot
    (earliest-start-time"2004-05-11[00:00-5]"default)
    (latest-finish-time"2004-05-11[23:59-5]"default)
    (status"possible"))
  (time-slot
    (earliest-start-time"2004-05-13[00:00-5]"default)
    (latest-finish-time"2004-05-13[23:59-5]"default)
    (status"possible")))
 :
```

Application II: Extracting Simple Event Structures from Newswire Texts

- Input: newswire texts (WSJ collection in Treebank)
- Output: event tuples

"*A form of asbestos once used to make Kent cigarette filters has caused a high percentage of cancer deaths among a group of workers exposed to it* ***more than 30 years ago***"

event ID

```
e20352416: (_, make/INF/None, _, filters)
e20402264: (form, used/None/None/PAS, _, once, (e20352416))
e20426800: (workers, exposed/None/None/PAS, {_-|>30.0_year|}, to)
e20429440: (form, caused/PRES/PERF, _, percentage)
```

subject predicate tense aspect time participant

Extracting Event Structures

- Identify temporal expressions using a statistical classifier
  - Recover **-TMP** function tag
  - Training a Support Vector Machine for each of ADVP, NP, PP, and SBAR on WSJ sections 2-6
  - Testing on WSJ sections 7-8

"*...workers exposed to it* ***more than 30 years ago***"

```
(NP
  (NP (NNS workers) )
  (RRC
    (VP (VBN exposed)
      (NP (-NONE- *) )
      (PP-CLR (TO to)
        (NP (PRP it) ))
      (ADVP-TMP
        (NP
          (QP (RBR more) (IN than) (CD 30) )
          (NNS years) )
        (IN ago) ))))
```

Simple Event Structures in Newswire Texts

"*After the shuttle Challenger <u>exploded</u>* ***17 years ago*** *, the shuttle fleet was <u>grounded</u>* ***for two years*** *as NASA <u>investigated</u> the mishap .*"

(Publication time: 21:23, 2003/02/02 EST)

```
e1: (shuttle, exploded/PAST/None, ?t1:={_-|17.0_year|})
e2: (NASA, investigated/PAST/None, ?t2:=_, mishap)
e3: (fleet, grounded/PAST/None/PAS, (?t3:=[_:+|2.0_year|]) > ?t1)
?t3 = ?t2
```

```
t1 = 1986.0
t2 = t3 = [1986.0_year:+|2.0_year|]
```

Conclusions

- Temporal scenarios in Natural Language are modeled as Temporal Constraint Satisfaction Problems.
- Human calendars are modeled as constraint systems.
- A typed language *Time Calculus* is designed to capture the meaning of temporal expressions and other temporal information.

The Future?