TERQAS Workshop Session 2
March 11 – 15, 2002

Table of Contents:

ATTENDEES 1

AGENDA OF THE WORKSHOP 1

GENERAL SESSIONS AND PRESENTATIONS 2

- SESSION OPENER 2
- PRESENTATION: *ANNOTATING TEMPORAL RELATIONS* (GRAHAM KATZ) 2
- SESSION ON CORPUS BUILDING (DRAGO RADEV) 3
- PRESENTATION: *ANNOTATING EVENTS AND TEMPORAL INFORMATION* (ROB GAIZAUSKAS) 4
- SESSION ON ONTOLOGIES (INDERJEET MANI) 4

WORKING GROUP PARALLEL SESSIONS 5

- ONTOLOGY/TENSEML WG 5
- QUERY/CORPUS WG 6

TASKS AND MILESTONES FOR SESSION 3 (APRIL) 8

- ACCOMPLISHMENTS: 8
- TASKS 9

Attendees

James Pustejovsky, Michael Bukatin, José Castaño, David Day, Lisa Ferro, John Frank, Rob Gaizauskas, Robert Ingria, Graham Katz, Inderjeet Mani, Mark Maybury, Bev Nunan, Drago Radev, Erich Rauch, Antonio Sanfilippo, Roser Saurí, Beth Sundheim, Marc Verhagen.

Agenda of the Workshop

March 11	morning	Session Opener (James Pustejovsky) Presentation: *Annotating Temporal Relations* (Graham Katz) Session on Query Corpus Building (conducted by Drago Radev)
	afternoon	Working Group parallel sessions Plenary Session
March 12	morning	Presentation: *Annotating Events and Temporal Information* (Rob Gaizauskas) Session on Ontologies (conducted by Inderjeet Mani)
	afternoon	Working Group parallel sessions

March 13	morning	Working Group parallel sessions
	afternoon	Plenary Session
March 14	morning	Working Group parallel sessions
	afternoon	Plenary Session
March 15	morning	Query/Corpus WG session Closing Session: assignments of action items and wrap-up

General Sessions and Presentations

Session Opener

To make the workshop more feasible, it is agreed that the Working Groups (WG) will be merged according to the goals defining each one of them. They become restructured as follows:

1. Query/Corpus WG, integrating:
   - the Corpus Collection/Definition WG
   - the Query Corpus Construction WG

2. Ontology/TenseML WG, constituted of:
   - the Ontology WG
   - the TenseML WG
   - the Algorithm WG (which is dependent on the previous two)

3. Evaluation WG

Presentation: *Annotating Temporal Relations* (Graham Katz)

Presentation:

The project aims at annotating the relations between events and states in complex sentences. The annotation used is defined so that (a) it is as neutral and language independent as possible; (b) it allows to distinguish the mark-up from the interpretation; and (c) it can convey both structural/morphological info and semantic info. Different annotations can be related as being equivalent, consistent or maintaining a subsumption relation.

The final purpose is to provide with lexical semantics information (structural restrictions and lexical preferences; e.g., met > marriage?). The corpus is constituted of around 200 sentences from 11 different languages. There is a checking of inter-annotation agreement.

Issues arisen from the presentation wrt TenseML:

- Some data in this project (e.g., verbs that can have both an event or habituality interpretation) show the necessity of using underspecification, deferring the decision until more information becomes available.
- It seems also important to annotate temporal-related nominalizations (e.g., *employee*, *graduated*, etc.)
- Confidence measures may help in avoiding excessive annotation.
- It would be interesting to know to what degree the temporal ordering of events based on tense interacts with lexical preferences. This goes very much along the same lines of the project Inderjeet Mani talk about at the previous workshop.
- It would be interesting to have some sort of mark-up for modality, intensionality, etc. Also, to mark the pervasivity of events (relevant with verbs like *prevent*, *cancel*, etc.).

Session on Corpus Building (Drago Radev)

The Query Corpus:

A good starting point seems to define the parameters that may be relevant for the kind of queries we want to collect. For instance:

1. The linguistics of questions and text
   - generated from a single sentence
   - from multiple sentences
   - from multiple documents
   - using general knowledge
   - hypothetical or modal questions
   - …
2. What kind of person that would ask each question
3. The frequency of each kind of questions (on the view of the WS task)
4. etc.

The WG can start examining existing typologies (e.g., Anderson/Belknap - taxonomy of questions, mid 1970s) and then building a typology of time-related questions.

Possible sources of questions:
- subset consistent with excite/TREC/etc.
- corpora: TREC/Excite/Encarta/TREC-Michigan/trivia.net

The Document Corpus:

Initial proposal:
- 100 docs. from ACE
- 100 docs. from DUC
- 100 docs. from Propbank (WSJ) (alpha version released last December)
- 100 docs from Reuters 21578 (from Inderjeet)

Presentation: *Annotating Events and Temporal Information* (Rob Gaizauskas)

Presentation:

The project introduced here has been developed in the framework of Andrea Setzer PhD research. The project focused at the annotation of:

- Events: verbs or verb groups
- Time expressions

Events are further specified according to the following attributes:

1. Tense
2. Class:
   a) occurrence (*crash*)
   b) perceptual (*see*)
   c) reporting (*say*)
   d) aspectual (*begin*)

On the other hand, time expressions are classified into simple (*last Tuesday*) and complex expressions (*16 seconds after the crash*).

The temporal relations that where annotated are those concerning an event and a temporal expression, or two different events.

The project also involved the development of a closure system able to infer the complete set of possible relations among events and time expressions for any given text.

Related documentation: Andrea Stezer's thesis, available at:

http://www.andrea-setzer.org.uk/PAPERS/thesis.ps
http://www.andrea-setzer.org.uk/PAPERS/thesis.pdf

Other papers on her temporal annotation work:

http://www.andrea-setzer.org.uk/research.html

Issues arisen here:

Starting from this annotation system, it would be useful to also encode:

- relations with multiple events
- anchoring not only with tensed events, but with untensed ones, nominals, etc.
- apposition sequences
- relations between 2 different times (e.g., *Monday before noon*)

Session on Ontologies (Inderjeet Mani)

Ontologies that don't appear to be really relevant wrt our particular task are excluded from consideration. From this point of view, SIMPLE is ruled out since the English version doesn't meet our particular requirements.

There is a discussion on what aspects should our ideal ontology cover. The debate takes SUMO (built out of several other ontologies) as a reference point.

Issues discussed:

- What support do we need for granularity? (e.g., *Tuesday, Tuesday afternoon*,…)
- If we want to coerce local references into temporal references, we also need an ontology of space information.
- About points and intervals: it is important to be able to distinguish points within intervals (e.g., *the starting of a meeting*)
- Also about intervals: sometimes parts of intervals can be seen as referring to intervals (e.g., *a meeting on Tuesday* vs. *a meeting starting Tuesday at 10pm*). There is agreement in that the best here is to keep the annotation as simple as possible, unless there is evidence for subeventive structure.
- The ontology distinctions should be driven by the kind of questions we want to be able to answer.
- What set of temporal reasoning tools do we want to have? We can borrow info from different approaches, like Event Calculus or Event DB Mangement.

Working Group parallel sessions

Ontology/TenseML WG

Here follows a summary of the main issues discussed by the WG:

1. It is desirable that TenseML can cope with:
   - modality (*the spokesman believed they escaped yesterday*)
   - conditional reasoning (*yesterday, according to the spokesman, ...*); those cases could be considered as reporting verbs, in the line of Sabine Bergler Ph.D. Thesis.
   - counterfactuals
   - negation

2. TenseML has to be defined also wrt structural ambiguity --in particular, when the context doesn't help at all. What to do in those cases?

3. There is agreement in that the more theory-neutral TenseML is, the better.

4. There is also agreement in that if we don't want to rely too much on the power of the inference system, we need a fine-grained tagging (to deal with cases such as *yesterday before noon* vs. *yesterday before Christmas*).

5. Anchoring needs 2 different levels:
   - Internal: local
   - External: ISO anchoring

6. It is necessary to have a mechanism for computing the relations between different kinds of temporal id's, especially in cases of delayed evaluation (that is, when we don't know yet the

actual time in which a particular event is anchored, such as the date of the text). Something along the lines of a lambda function. E.g., *last week* => `anchor(2) = anchor(3) – 1`

7. Currently, the Sheffield Annotation Schema doesn't handle the relation between 2 different times (e.g., *Monday before noon*). Instead, it only encodes relations between times and events (e.g., *Monday before the crash*). TenseML should be allowed to annotate both kinds of relations.

8. It could be interesting to have a library of relations between event types (e.g., *crash - die*), which would provide with info for some sort of closure.

9. Information not currently annotated by the Sheffield Annotation Schema but that TenseML should encode:
   - Identity of different relations
   - States: we want to annotate those that are updated in the time frame of the story being narrated (stage level states).
   - Tense: we also want to annotate non-tensed verbs (`tense = none`).
   - Aspect: we want to use such info as signal. E.g.:

at	*noon*	*he*	*had*	*left*	(= he had left before noon)
	`<signal>`		`<signal>`	`<event>`	
	`tid = 1`		`tid = 2`	`RelId = 2`	
				`RelType = before`	

The same can be done with progressive, which expresses interval of time.

Query/Corpus WG

The WG main tasks were:

1. Concerning the Query Corpus building:

1.1 Manual generation of questions from sample texts:

The idea behind the task was to have a first contact with temporal related questions, and create an initial set of parameters that would allow their classification. The initial set of sample queries has been generated from 3 articles. The 3 articles are available at:

http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/QueryWG-text1.txt
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/QueryWG-text2.txt
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/QueryWG-text3.txt

Some of the criteria followed when generating the set of questions were:

- trying to avoid yes/no questions
- look for a varied syntax
- think on possible hidden information (from other documents)
- counterfactual questions (*When was gambling banned in Las Vegas*)
- hypothetical questions
- think on questions for which the answer has part of the info in text - the rest is not

- questions that require comparison
- questions involving intervals
- questions with a not too obvious answer
- interpolation and extrapolation

These queries have been classified according to the following parameters:

1. Syntactic answer type: yes/no question vs. short-phrase answer vs. narrative answer
2. Syntactic question type
3. Strategy to find the answer: question can be answered purely based on the text, less than 100% confidence, etc.
4. Number of different answers
5. Time frame: point in time, multiple points in time, frequency, interval, always, never
6. Whether the question is fully specified
7. Number of related events: 0, 1, 2
8. User type (also related to the domain of the text/question)
9. Amount of background knowledge required to answer

For a collection of the queries generated (some of them already classified), visit also:
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/QueryWG-questions.txt

As a result of this exercise, it's been concluded that Reuters documents have appeared to be too difficult wrt TERQAS task. Agreement in dropping them from the previous selection.

1.2 Selection of a set of temporal related questions from already available corpus of questions.

The source corpus is a sample of 4000 questions from Excite. The selection of temporal questions aims at the creation of a taxonomy adequate to classify temporally relevant questions. After working with the initial several hundred questions, the WG has come out with a temptative set of classification parameters. The "Draft Template for Temporal Question Taxonomy" (by Patrick and Beth) is reported at:
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/QueryWG-queryTaxon.txt

2. Concerning the creation of a document corpus

2.1 Selection of the texts that will constitute the corpus

It's been accorded that the TenseML text corpus will be constituted of:

a. DUC (100 articles): Training (50), Development (25), Testing (25)
b. ACE (100 articles): Training (50), Development (25), Testing (25)
c. PropBank (??? articles)
d. TBD, either AP News or NAMTC (North American News Corpus) (100 articles):

2.2 Adoption and/or Development of a set of corpora tools

A KWIC tool has been adapted. Tempex has been run over articles and results have been concordanced. ACE training set has been indexed and is now available on MITRE machines.

Tasks and Milestones for Session 3 (April)

1. Accomplishments:

1. The spec language TenseML is now designated as TML, to stand for **Time and Event Markup Language for Natural Language**. Focus is on the needs of the Natural Language community.
2. The Ontology WG has completed its assignments and is converging with TimeML.
3. Corpus Collection is well underway. All four subcorpora are available on the MITRE machines, except for PropBank. The four annotation subcorpora are:

   a. DUC (100 articles): Training (50), Development (25), Testing (25)
   b. ACE (100 articles): Training (50), Development (25), Testing (25)
   c. PropBank: (?? Articles):
   d. TBD, either AP News or NAMTC (North American News Corpus) (100 articles):

   Background/reference corpora are being assembled for study, training, statistical analysis, etc. These include:

   a. BNC (not yet)
   b. AP 92-96 (subject to permission)
   c. NAMTC (in house)
   d. TDT2 (in house)
   e. Penn TreeBank (in house)
   f. TREC 4 and TREC 5 (in house)
   g. TIDES-MiTAP (Promed and disease related articles) (in house)
   h. Reuters 21578 corpus (in house)
   i. Medline (at Brandeis)
   j. Enthusiast (Dialog corpus) (in house)
4. Existing query corpora are being acquired. Methodology is being developed for how to elicit new questions.

   a. Q&A TREC 9 +10 questions (<1500): TREC 9 are better than TREC 8 questions.
   b. Excite Log of user input (209K unfiltered):
   c. The Marc Light Corpus (TMLC), (>1.2 M)
   d. Invented Questions (<50)
   e. Encarta Questions (??)
   f. Riloff's corpus of questions (not yet)

   Extraction of time related questions and classification from the above corpora is in progress. WG is preparing a minimal set of attributes and possible values for the classification.
5. Concordances created over corpora:

   a. DUC: (Marc V.) Tempex was run over articles and he concordanced results.
   b. zKWIC: ACE training set has been indexed and is available on MITRE machines.
   c. Bran Boguraev's concordance results from Kick Off Meeting
6. Towards an initial specification for TML (version 0.1):

   a. Incorporating much of STAG (Sheffield Temporal Annotation Guidelines)
   b. Compliant (after study) with TIMEX2 guidelines.

c. Introduce Link tag: an object that links events/times to events/times.

d. Introduce a State tag: annotate only states that are updated in the context of the narrative being tagged. Any state persistent throughout the entire article would not be tagged as a state.

e. Enrich time relations: add immediately-after, immediately-before.

f. Introduce scale as a relation attribute: we need to convert preexisting Timex data into the TML standard

g. doa tag will have a tid attribute.

h. event identity:

i. Added "none" as a value for tense attribute

j. Aspect is labeled as a signal: (none, progressive, perfective)

k. Add temporal inline functions: (e.g., last week): doing temporal math. Track this as an enrichment over the TIMEX2 guidelines.

l. Possibly identify "Event Clusters" or "time frames". This would be useful for clustering related events in a narrative, temporal segmentation of the narrative.

m. Brief discussion of negation and modality. Use a polarity attribute on negative propositional content:

    a. *The plane did not crash.*
    b. N*o survivors were found.*

n. Enrich Event Typology:

o. Adopt the XML schema definitions rather than writing merely DTD.

p. Add hooks to the event ontology for event entailment operations.

q. Event and time closure operations as part of TML.

r. Alternative to annotating the verb as a signal to the event.

s. Hooks to init and cul attributes to events, or either reify init and cul as events.

    i. *The party will begin at noon.*
    ii. *The man began the lecture at noon.*

t. Discussion point (pending):
Should we make "doa" an event, such as the default utterance event, with type reporting event?

2. Tasks

1. Short Description of Sheffield Temporal Annotation Guideline (STAG) (RJPI/JDP)
2. Short Description of Hobbs/Semantic Web Proposal (per time) (SW) (RJPI/JDP)
3. Short Description of SUMO temporal types (AS)
4. Initial Spec Notes on TML as defined this week. (RJPI/JDP)
5. Short Comparison of STAG, SW, and SUMO. (AS)
6. Clarify the status of corpora, ensuring that we have properly partitioned development, training, and test corpora for each subcorpus. (BS/DR)
7. Make corpora available to workshop members over the web by password protection or network access to MITRE. (RS/BN)

8. Secure rights for use of British National Corpus (Hanks will give David Day and Bev contact information). (PH/RG/BN)
9. Secure permission for use of AP News corpus (Patrick Hanks will notify Bev.) (PH/BN)
10. Extend existing concordancing tools (Robert's zKWIC) and Marc's ad hoc tool. (BS/MV)
11. Secure license of LingoMotors' CPA tools (JDP)
12. Research Mike Scott's Wordsmith concordancing tools; (MV/PH)
13. Examine MiTAP corpus, particularly the ProMed subcorpus. (IM)
14. Refine and develop the parameters for typing and analyzing the question corpus. (Query WG)
15. Contact Riloff for corpus of questions and types. (IM)
16. Assemble expressions for zKWIC analysis by Beth for corpora at San Diego. (RS)
17. Look into the Steven Bird graphical markup language at Penn (RS/RG/GK)
18. How do we do more sophisticated temporal inferencing associated with scales? Cf. Fikes and Wu. Add this to reference list. (GK/RG/IM)
19. How do we do temporal arithmetic according to ISO standards? (RJPI/JDP/GK)
20. Examine the role and contribution of confidence in the markup. How would it be marked up? (ER/MB/JF)
21. Future development of Annotation Tool:

    a. STAG AAT extensions (RG/MV/LF)

    b. Alembic extensions (LF/DD/MV)
22. Have algorithms WG start thinking about extending Tempex, analyzing the preprocessing tools, and using parsing and finite state techniques for pulling out expressions. (ER/MB/JF/JC)
23. Excite Log (Query WG: Drago, Beth, Lisa, Mark, Patrick, Roser)

Notes available at:
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/Notes03.doc
http://www.cs.brandeis.edu/~jamesp/arda/time/documentation/Notes03.html