

LING 7800 COMPUTATIONAL LEXICAL SEMANTICS
SPRING 2016
PROFESSOR MARTHA PALMER

**QUALITY REVIEW OF THYME EVENT ANNOTATION
AND
A REVIEW OF AUTOMATED EVENT-ANNOTATION METHODOLOGIES**

March 19, 2016

Matthew D Thomas
University of Colorado at Boulder

TERM PROJECT PROPOSAL

The goal of the *Temporal Histories of Your Medical Events* (THYME) project is to develop a human-annotated corpus of medical records annotated with temporal relations between events, occurrences, states, dates, and procedures.[2].

Two annotation passes were carried out on the THYME corpus. The (primary) temporal annotation pass focuses entirely on temporal annotations and permits, for example, "*the same heart attack [to be labeled] as an EVENT ten different times in the document*"[2] whereas the purpose of the (secondary) coreference annotation pass is to figure out which specific mentions of events and entities are the same.[3]

The underlying goal of this term project is to understand of temporal structure of the THYME corpus by accomplishing the following:

1. Explain the two annotation methodologies and their challenges
2. Synchronize the results of the two annotation passes
3. Focus on identifying and quality-checking annotations of CONTAIN relations. For example, "*CONTAIN MRI Tumor, where what the MRI really contains is the observation of the tumor. The tumor itself existed before the occurrence of the MRI, and continues to exist afterwards.*"
4. Describe automated approaches to event annotation
5. Discuss the impact that annotation quality might have on automatic classification

The quality check component will attempt to provide examples of inconsistencies that could be fixed via hand-correction and determine whether some inconsistencies could be fixed deterministically. This project will not attempt to implement the suggested fixes. This work will hopefully serve to suggest methods of cleaning up the two annotations which in turn could improve results of classifiers "able to construct an accurate timeline of when the events in a given medical record occur relative to any fixed dates present and relative to one another"[2].

0.1 NOTES FROM READING

0.1.1 [2]

"Note also that although we are putting these on a timeline, you can label many EVENT entries which refer to the same actual "event" on a timeline. You can (and should) label the same heart attack as an EVENT ten different times in the document, so long as it's a separate event in each. We have a secondary "coreference" annotation project which will figure out which events are the same, and therefore we want to capture every EVENT regardless of its previous mentions." [2]

0.1.2 [3]

asdf...

0.2 MEETING NOTES

Tim will send me all the data today. It's in Standoff annotations, xml files with spans, and the span means... you just count characters within the raw files... so first step is to get an understanding of how the data is represented.

What are the questions for the quality-check hypotheses. Look for if there are things called entity in the entity annotation that are called events in the event annotation. Temporal pass = event annotation Coreferent/entity pass = annotate both events and entities

There is a mundane task of merging them in terms of spans and things like that. Tim has already started doing this...

The more interesting task is having a bunch of events with relations which are marked of being coreferent... things in the future may be marked with things in the past.... or real things with hypothetical things.

These inconsistencies would totally screw up the automated classifiers. I can say that the guidelines were inconsistent... so I need a thorough enough understanding of the guidelines to write about why the annotation was done in different ways for each pass.

When I do the presentation to the class, want to be able to tell the story about where the inconsistencies are coming from, why they are there.

Implications: 1. How can guidelines be created for separate annotation layers to be maintain consistency. 2. How badly this messes up the machine learning. 3. Can some of these inconsistencies be fixed non/deterministically without having to re-annotatet everything from scratch. a. By hand correction? b. Deterministically? Have a few examples of things that could be fixed via hand-correction and examples of things that could be fixed deterministically... but I wouldn't have to actually make the fix.

0.3 ToDo

- DONE Meet with Kristen (from video, lead annotator) and Tim to clarify other quality check hypotheses.

0.4 PROJECT PRESENTATION

Give a 10-15 minute presentation on the project.

0.5 INTRODUCTION

You are expected to turn in a 5-10 page, single spaced paper describing your project, and give a 10-15 minute presentation on it.

The goal behind this assignment is to have you become familiar with a specific research area and take a stab at moving the state of the art forward. The project could be primarily linguistic analysis or primarily a description of algorithms or something in between.

You should assume that you will have to read something like 3 or 4 papers over and above the class required readings to ground yourself in the research area. You will then define an experiment or a set of analyses or a system that you will run, perform or implement, respectively, to explore some aspect of your research area. Think in terms of defining a proposal for a major new research project. You only have to do enough preliminary new research to clearly define your approach, or to justify it. Your justification could potentially also be in the form of results from another area, such as psycholinguistics, that provides compelling evidence for why your approach could be better than the current state of the art.

Become state of the art results in a topic, and figure out where you'd want to go next. Might mean a re-implementation and improvement of some else's work, and just an error analysis of the results and discussion of how it might be improved.

Do a project that is hypothetically publishable.

0.6 TOPICS FOR MEETING WITH PROFESSOR MARTHA PALMER

1. Finishing up SPM project so that I can focus the rest of semester on this project.
2. I'd like to try to finish the project by the end of March.
3. Professional MS in NLP
 - Possibility of writing a letter regarding partial completion.
 - Class she is teaching in the Fall, could be taken remote?
 - Counts for CS?

4. Missing class on March 17th.

0.7 TERM PROJECT IDEAS

My interest areas

- Automatic Ontology creation
- Automatic code generation given project proposal
- Dialogue Agent
- Conversation summarizer
- Something to do with Event-Structure?
- Word2Vec... perhaps run lemmatizer first? What about Semantic Vectors? Syntactic Vectors... etc. What about Word2Vec for morphemes to do morphological analysis?
- Automatic sense grouping using Word2Vec? Somehow break words into different senses depending if there are very disparate groups of words... this would have to look at the Word2Vec for the related words and compare them. Maybe using an ML clustering algorithm?
- "Great term project: Research the Semantic Web or Google Knowledge Graph and give a presentation of it to the class."
- "Another Potential Term Project: System called Never-Ending Language Learning (NELL) is more or less unsupervised induction of facts, looking for relationships."

Projects ideas mentioned in class

- another program "Communicating with Computers"
- THYME: Temporal History of Your Medical Events
- CLEAR earth: suite of NLP tools and port is to the GL sciences domain
- Patient-focused Project
- EPIC: NSF funding with NCAR/UCAR uses Hurricane data
- Experiment to see if new data set, train things with -Goal instead of arg2 and argM, gets better results?
- Go to the aclweb.org website and look at the papers from the ACL events, hundreds, which can be used as a springboard for your project/paper

- Another Term Project: Lynn Shubert has a good NLP system that produces this kinds of logical representations for sentences. He has been using that to extract logical facts from sentences. It works better with smaller sentences. It is the preliminary version of a system that could be used to build up Ontologies/Axioms of the world.

Project ideas from previous instances of class

- TBD

Final Project Ideas from Term Project Document Google: preposition super sense tagging, Hindi/Chinese/Korean, compare PB function tags to Preposition super senses My Science Tutor, NIH THYME, NSF ClearEarth: 1) comparing dependency-based SRL and phrase structure SRL; 2) train SRL on ArgX-Tag rather than ArgN and ArgM (Skatje?) DARPA CwC and DTRA - eTASC: 1) simplifying VN wh- syntactic frames; 2) GL-VN mappings; 3) CCG-VN parser

DARPA DEFT and NIH: For THYME/RED, synchronize coref IDENT entity chains with temporal markables (different spans) and check for modality violations and replace with bridging, also look for events in entity coref chains - are they implicit events in whole/part relations? Add an event/subevent tag

DARPA DEFT: RED → analytics for joint annotation task for RED, compare ACE/RED and ERE/RED, similarities and differences

DARPA DEFT AMR: Discourse Frame files for PropBank (English, Arabic, Hindi or Chinese) English: Run Jeff's aligner, extracting PDTB discourse connectives, map them to AMR concepts Create PB style Frame Files that codify current practice

DARPA Lorelei: Projecting English semantic role labels onto another language (Persian?) via Giza++ word alignments and magic → see Ghazaleh Kazeminejad

Or topics based on recent ACL papers:

Fast and Robust Neural Network Joint Models for Statistical Machine Translation <http://acl2014.org/acl2014/1/pdf/P14-1129.pdf>

Linguistic Structured Sparsity in Text Categorization. Dani Yogatama and Noah A. Smith. <http://www.cs.cmu.edu/~nasmith/papers/yogatama+smith.acl14.pdf>

Karl Moritz Hermann; Dipanjan Das; Jason Weston; Kuzman Ganchev
Semantic Frame Identification with Distributed Word Representations

Denis Paperno; Nghia The Pham; Marco Baroni

A practical and linguistically-motivated approach to compositional distributional semantics <http://aclweb.org/anthology/P14-1009>

Nal Kalchbrenner; Edward Grefenstette; Phil Blunsom

A Convolutional Neural Network for Modelling Sentences <http://aclweb.org/anthology/P14-1062>

Socher et al. Grounded Compositional Semantics for Finding and Describing Images with Sentences <http://www.aclweb.org/anthology/Q/Q14/Q14-1017.pdf>

Alona Fyshe; Partha P. Talukdar; Brian Murphy; Tom M. Mitchell Interpretable Semantic Vectors from a Joint Model of Brain- and Text- Based Meaning <http://www.aclweb.org/anthology/P/P14/P14-1046.pdf>

Low-Rank Tensors for Scoring Dependency Structures

<http://people.csail.mit.edu/tommi/papers/Lei-ACL14.pdf>

0.8 FIRST SECTION

"Really cool quote"

-Author

asdf

Bibliography

- [1] clear.colorado.edu, *End-to-End Systems: Temporal History of Your Medical Events (THYME)*, <http://clear.colorado.edu/compsem/index.php?page=endendsystems&sub=temporal>, [Accessed March 2016]
- [2] clear.colorado.edu, *THYME Guidelines*, <http://clear.colorado.edu/compsem/documents/THYME%20Guidelines.pdf>, Will Styler, Guergana Savova, Martha Palmer, James Pustejovsky, Tim O’Gorman, and Piet C. de Groen, Updated February 28, 2014, [Accessed March 2016]
- [3] clear.colorado.edu, *Clinical Coreference Annotation Guidelines (with excerpts from ODIE guidelines and modified for SHARPN/THYME)*, http://clear.colorado.edu/compsem/documents/coreference_guidelines.pdf, Arrick Lanfranchi, Kevin Crooks, and Mariah Hamang, Updated November 2, 2013, [Accessed March 2016]
- [4] clear.colorado.edu, *THYME Wiki*, https://clear.colorado.edu/TemporalWiki/index.php/Main_Page, Will Styler, Guergana Savova, Martha Palmer, James Pustejovsky, Tim O’Gorman, and Piet C. de Groen, Updated March 8, 2016, [Accessed March 2016]