**Term Project Ideas – Intention**

The goal behind this is 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 programming or could combine the two. You are welcome to work alone or in 2 person teams. If you do work in a team it should be interdisciplinary, it should include students from two different departments. 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. 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. You could also do serious comparison of two or three approaches at a detailed level, in which case you would turn in a longer paper of up to 15 pages.

In addition to your own Term Project, you are expected to be a discussant on another project. The Discussant assignments will be posted on the web page once the term projects are finalized. That will involve reading the project background paper(s), asking constructive questions during the project presentation, and turning in this [questionnaire.](http://verbs.colorado.edu/~mpalmer/Ling7800/DiscussantQuestions.pdf)

Past Projects

* Look at the web pages for previous versions of this course to see what other students have done.

**Term Project Ideas:**

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

RED – compare ACE/RED and ERE/RED, similarities and differences - Ali

Tense and Aspect annotation – look at PB annotation, compare to TimeML on same data, write Tense and Aspect annotation guidelines and annotate some data

Projecting English semantic role labels onto another language (Persian?) via Giza++ word alignments and magic – see Tim O’Gorman

GL-VN mapping – Michael?

A biomedical variant of GL –VN mapping suggested by Kevin Cohen:

In the biology literature, there are a lot of Change Of State verbs--phosphorylate, methylate, acetylate, etc.

From a domain-specific perspective, these are interesting because one way to think of a lot of biology is a series of changes in state of molecules, and that's what these verbs are about.

Kevin would be interested in looking at similarities between these COS verbs and differences between these verbs and non-COS verbs.  A theoretical question that you could ask is: do all of these COS verbs have their differences in meaning in the same locus of the semantic representation?  E.g., is it all about the qualia structures?  If so, is it all the same aspect of the qualia?  And then, when you compare them with the non-COS verbs, is the difference in meaning encoded elsewhere?

Then, the next question: these changes of state are sometimes reversible--e.g., you can methylate DNA, and you can also demethylate DNA.  Where is the locus of difference between these pairs of verbs?

A practical question would then be: can you make use of what you know about the relationships and differences between the meanings to scale up production of a lexical resource--e.g., by inheritance?

Some details: Kevin looked in VerbNet for some of these, and they're not well represented, but he did find some related verbs--hydrogenate, fluoridate, chlorinate; they're not the same, in that they're inorganic reactions (to the best of my knowledge), but related.

Kevin could contribute a corpus, a search tool, domain expertise, writing the paper, and some of the time that would be needed to do the mapping.

Something to do with PropBank, semantic role labeling or lexical semantics, for English, Hindi, Arabic, Chinese, some other language

Or topics based on recent ACL papers:

Fast and Robust Neural Network Joint Models for Statistical Machine Translation

<http://acl2014.org/acl2014/P14-1/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](http://www.cs.cmu.edu/~nasmith/papers/yogatama%2Bsmith.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*](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>

Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space (accepted long paper for EMNLP 2014)

<https://people.cs.umass.edu/~arvind/emnlp2014.pdf>