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 a description of algorithms or something in between. You are welcome to work alone or in 2 person teams. 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. 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.

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 a form that will be provided.

Past Projects

- Look at the web pages for previous versions of LING7800 to see what other students have done.

Term Project Ideas:

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/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

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

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