MINUTES: SUPERPARSER PROJECT VIRTUAL MEETING 21st JUNE 2002

1. Progress so far: Reports on Parsers, cache manager, and datasets.

   Julia and Steve have the CCG pipeline up and running and have a slightly positive self training baseline due toi the cleaning up of the dataset. They are in a position to start cotraining immediately.

   Rebecca and Miles have the Collins pipeline up and running with a slightly positive self training result (is this right? on the clean dataset?) and will start cotraining on the Beowulf Wednesay 26th

   LTAG has no new results.

   Cache mangaement is under development. Paul needs sample output trees from all parsers.

2. Discussion: We decided (again) against a common output format for all parsers. Parsers will output the output they output, labelled as to type. Pairs of parsers who agree to cotrain will write their own translation routines for the relevant types and ignore labeled data of other types.

   Annoop will investigate the uselfulness of adding functional tags to parser output.

   Mark will mail the list on Monday regarding the mass login to beowulf on Wednesday (he screwed up, but Miles did something.)

   Steve suggested everyone (along with cotraining etc.) simply parse as much data as possible in order to do an experiment on the one-iteration large retraining set condition of cotraining.

   Mark stressed the importance of getting cotraining working as opposed to establishing self-training baselines etc, which if necessary can wait.

   Mark reminded everyone of the comittment to do one reality-testing iteration of the cotraining loop with the Brown corpus in preparation for cross genre porting experiment

3. Actions:

   Everyone: UPDATE YOUR TASKS ON THE WIKI ACCORDING TO THIS MINUTE, sending me corrections if necessary.

   Paul: Complete implementation of the beam-search lab for class 11th July, debug with Steve and road test; Complete cache management module (with Steve); Bug people for output if/when they fail to send it.

   Rebecca: Post summary of the cache management discussion to Wiccie. selftrain on Julia's output and cleaned up data Start cotraining 26th June (with Miles) see above). Do Brown baseline. Parse as much stuff as possible.

   Anoop: Make I/O stages of LTAG compatible with Collins. Redo cotraining ewxperiment with the new clean data. Explore addition of functional tags. Do Brown baseline. Parse as much data as possible.

Stephen: Tell everyone where everything is at Edinburgh filespace. Calibrate costs of retraining on small datasets. Finish Supertagger. Parse lots of unlabelled. Do Brown baseline. Leave Edin 5th July.

Julia: investigating effects of iterative training with sets of increading complexity. NN selector description to Wicky. Explore regression for selection. Leave Big E 28th.

Miles: Explore other methods for regression for selection. Organize Beowulf accounts. Help Rebecca. Reduce the 3 billion features of ME to something reasonable. Explore reranking. Do Collins cotrain. Parse a lot of stuff. Brown baseline. Leave E 4th July

Mark: Prepare presentations on the project. There is a public meeting 15th at which we present the nature of the project and the preliminary work—I'll need to call on you all for stuff to make into slides and you should all plan to be there to answer questions from the funders. Maybe some of youse will help with the presentation itself. Leave Ed 4th July.