

File 20100513.1008: Notes from meeting with Tom Marso this morning:

I found Tom at his desk after the 5.0 certification telecon. He showed me the new machine ‘faraday’ which is not so loud in operation as Eric Richardson makes out to be. Approximately half the disk slots are active, as shown by green lights on the front of the rack unit. Eric is going to house it in a rack in the data centre where the machine will have clean power and air.

Tom and I met in the TISS Conference room until we were chased out by the planning meeting. It was long enough to go through Tom’s design and talk a lot about the problem of handling variant word forms based on root words. We discussed how to search word forms such as ‘walking’, ‘walked’, ‘walks’ from ‘walk’ by annotating the root word in the corpus as dirty and thereby implicitly considering every longer word below it in the tree as having virtually the same tag. We discussed how to handle the standard case of ‘secretary’ being descended from ‘secret’ and decided that it would be helpful to tag ‘secretary’ as clean which would apply only to ‘secretary’ when walking the tree, but it would not automatically un-dirty ‘secretaryness’ further down in the tree, which if not marked clean would be treated as implicitly dirty. Tom took my diagram with him and said it was enough of an idea to keep him busy for a week. I thanked him for his time and we talked about my school work for a little while. Tom wanted to be added to my Oxford WAR mailing list, which I have done.

Another interesting thing that came out of our discussion was the idea of tracking grams that do not appear in the corpus. This is especially interesting to me because it’s a discovery that came from empirical work, it was not thought of as important *a priori*.

## References