Corpus Navigator Status

Jonathan Clark

Monday, May 6, 2008

1 Project Issues

1. Paul has begun sending me weekly updates of his progress through the Urdu EC
2. Aside: What time can we meet for the Grammar Formalisms Demo on Friday?

2 Open Issues

1. Can we simulated factored translation in Avenue? Or do we have to use feature constraints to get the same benefits?
2. What lexical resources do we plan to use for our Avenue experiments?
3. How do we prevent overfitting? (Should I use machine learning-type optimization anyway?)
4. We don't consider probability mass of language covered

3 Research Plan

1. Create navigation "simulator" that simulates navigation without interaction from a real user (all scoring functions and knowledge sources are dummy stubs, but we can request the actions: translate, align new, or align with suggestion) *By MON 5/2*
2. Decide Opening Book *By MON 6/9*
3. Integrate ParaMor (add external API to ParaMor and associate our morphemes with Christian's paradigms) *By MON 6/9*
4. Build an external morphological analyzer that segments with ParaMor and then assigns our features *By FRI 6/13*
5. Implement heuristic evaluator with lookahead *By MON 6/30*
6. Build MILES Implicational Universal Knoweldgebase ("If you know there's animacy marking, what else do you expect to be marked?") *On WED 6/2*
7. Build MILES "Related Features" Knowledgebase ("If you know there's animacy marking, what do you expect might interact with animacy?") *On WED 6/2*
8. Get Finsihed Urdu Gold Standard from Paul *By 6/30*
9. EXPERIMENT: Random Selection of Sentences vs Navigation of sentences (How many new feature constraints / morpheme-annotation pairings do we learn after each sentence?) – This shows that navigation can learn (almost) as much from less data *By FRI 7/4*
10. Prepare Urdu Parallel Training and Testing Data *By FRI 7/11*
11. EXPERIMENT: Moses with and without our annotated morphology "factors" – This shows the data from Feature Detection is useful to SMT *By MON 7/14*
12. Implement Feature Constraint Generator *By MON 7/28*
13. EXPERIMENT: Avenue Rule Learner trained on whole EC and then selected portion of EC (includes our annotated morphology factors AND our suggestions for feature constraints; both experiments use external lexical resources) – This shows Feature Detection is useful to Statistical XFER while using less data *By FRI 8/8*
14. Build WALS-to-MILES Feature Mapper *By MON 8/18*

4 Upcoming Dates

1. 5/12 MON - (Next Meeting) Lori will be out of town
2. 5/19 MON - (Next+ Meeting) Jon will return from California in the evening
3. 5/26 MON - (Next++ Meeting) Jon and Bob will be at LREC