Proposal Number: 1320684

Panel Summary: Panel Summary A brief statement of what the proposal is about: The core idea of this proposal is to develop a methodology for the discovery and use of systematic linguistic inferences identified with adjectives. This will require the development of an inferential model for adjectival semantics, connecting the model to data mined from the web, and revising the model based on human judgements via Mechanical Turk. Intellectual merit: - Strengths: The panel felt very strongly that this line of work toward better understanding of adjectival semantics is important and much needed. Further, the tight focus on scalar, predicative, and intensional adjectives was seen as a strength, making real progress on theoretical and corpus-based linguistic analysis possible. For example, the proposed work could have significant impact on the difficult problem of textual entailment. - Weaknesses: The way in which expert and lay annotations would interact was not at all clear. The claim is that "in the wild" annotations would be used to refine the inferential models, but the mechanism was not obvious to anyone on the panel who reviewed the proposal. It seemed that the structure-to-inference mappings would apply only to intensional adjectives, despite earlier indications that the model would be more universal (at least in the context of the adjectives under consideration). Further, only scalar adjectives have a clearly defined extrinsic evaluation (textual entailment), but too few details are provided on how the mappings would be used in specific tasks. Though the pattern-based approaches reported in the literature that will be leveraged here to find examples of (scalar) adjectives show promise, they produce tremendous amounts of junk when applied to the web and the frequency of sought after examples of the patterns is small. This could pose significant problems. Related to this, the panel thought that FactBank should be used to explore predicative adjectives with clausal complements. Of great concern was the fact that consultants will be paid from grant funds but there is no description of what they will be doing. - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts? The inferential models for the semantic analysis of the three classes of adjectives is the key innovative component. However, it is not entirely clear how relevant or useful these models will be in extrinsic tasks. Broader impacts, including enhancing diversity and integrating research and education: - Strengths: The PIs will help educate graduate and undergraduate students. - Weaknesses: There are no details on education or outreach in the proposal body. This is a serious weakness. The students seem to be relegated to implementing and overseeing annotation tasks. A more compelling case could be made for the utility of the models in other NLP tasks. - Adequacy of the post-doc mentoring plan (if required): Not required. - Adequacy of data management plan: The plan is terse but was deemed adequate. - Results from prior NSF support (if applicable): Results from prior NSF support are quite good. Panel recommendation (check one): __ Highly Competitive __ Competitive X_ Low Competitive __ Not Recommended for Funding by the Panel Additional suggestions: Some panelists wanted more justification for why the specific adjectives classes were chosen, e.g., their frequency and coverage. Justification of rating, including key strengths and critical weaknesses: The proposed work is important and the PIs are highly qualified. The computational linguistics community has paid far too little attention to the semantic analysis of adjectives. However, the proposal does not make a compelling case for the utility of the inferential patterns, the roles of expert and crowd annotators are not clear, and the use of the web to mine patterns without significant attention to data quality seems a bit naive. There is far too little detail about education and outreach. The summary was read by the panel, and the panel concurred that the summary accurately reflects the panel discussion.

Panel Recommendation: Low Competitive

REVIEW #1:

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit. A brief statement of what the proposal is about: This proposal investigates adjectives. It considers three types: scalar adjectives, adjectives that take a clausal complement, and intensional adjectives. The goal is to determine what inferences these adjectives allow, and to formulate the results (or some of them) as templates of structure-to-inference mapping. A combination of expert and crowd-source judgments will be used. Intellectual merit: - Strengths: Interesting area (adjectives). The PIs are highly qualified for this work. - Weaknesses: + "Structure-to-inference mappings" only seem defined for intensional adjectives, despite what is announced up front. + Relation between predicative adjectives with clausal complements and FactBank not explored -- doesn't FactBank contain examples that could be studied? + It is unclear what the PIs expect to find in terms of differences in judgment between sophisticated judges and MTers, and how they will use these differences in their work, or what the community at large will learn from the double annotation. + Why are these three classes of adjectives chosen? - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts? The creation of a resource on adjectives could have the potential of truly helping other research in NLP. The actual proposed methods appear fairly straight forward. In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts. Broader impacts, including enhancing diversity and integrating research and education: - Strengths: The resulting insight will be broadly useful in NLP and computational linguistics. - Weaknesses: "Female and minority students will be recruited for providing Gold standard judgments" -- this is a rather restricted broader impact. - Adequacy of the post-doc mentoring plan (if required): n/a - Adequacy of data management plan: adequate Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable Summary Statement Additional suggestions: Justification of rating, including key strengths and critical weaknesses: While the overall goals of the project are laudable, the specifics are less enticing. The notion of structure-to-inference mappings is interesting, but only mentioned for one type of adjective. The investigation of adjectives with clausal complements is not seen in the larger context of similar inferences allowed by verbal elements. The promised investigation of the relation between expert judgments and crowd-sourced judgments remains only hinted at.

REVIEW #2:

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit. A brief statement of what the proposal is about: The goal of this research plan is to develop an inferential model for three types of adjectives in natural language. This model will be constructed by using a small amount of data annotated by humans with some level of linguistic training and then refined based on comparisons from annotations by native speakers with no special training in linguistics. Intellectual merit: - Strengths: Deciding textual entailment by automated approaches is a complex task. The inferential model the PIs intend to develop could lead to improvements in TE tasks. - Weaknesses: While the three classes of adjectives that will be the focus of the research seem to require the analysis proposed in this project, it is not justified properly why these three classes should be studied first. Related to this, how much to these classes cover? Can all adjectives be categorized along these three groups? The evaluation component of the research plan is weak, only the scalar class seems to have a clear extrinsic evaluation within the TE task. The evaluation of the mappings is described briefly, but little details are provided regarding the integration of the mappings on specific tasks. - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts? The innovative aspect relates to the inferential models for the semantic analysis of three classes of adjectives that the PIs want to develop. However, it is not very clear how relevant or useful these models would end up being. In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts. - Strengths: One of the PIs and the consultant are females, which could lead to a more successful recruitment of females students to the project. The research plan includes graduate and undergraduate students that will be involved in the annotations. - Weaknesses: It seems the involvement of the students in the project will be limited to the annotation task and while this task includes some analysis is not really clear if other research experience will be available to the graduate students. Other than the fact of having females in the senior personal and consultants, there is no specific description or focus on increasing diversity. - Adequacy of the post-doc mentoring plan (if required): N/A - Adequacy of data management plan: The data management plan is a very terse description of making "everything" available. Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable Summary Statement Additional suggestions: A stronger motivation for the need of the proposed inference patterns would be good as well as a more thorough extrinsic evaluation plan. Justification of rating, including key strengths and critical weaknesses: Inference in natural language is problematic to say the least and the research community in computational linguistics has not paid too much attention to the semantic analysis of adjectives. However, it is not clear that the proposed research will fill in the gap since the evidence showing the usefulness of the inference patterns that will be developed is weak. The use of the web to mine a representative sample of patterns in English seems intuitive, but also a bit naive. In particular, data cleansing, is barely discussed and it can be a major challenge according to how they plan to collect the data. The team is highly qualified to do the task. However, the involvement of the students seems limited to the implementation and overseeing of the annotation tasks.

REVIEW #3:

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit. A brief statement of what the proposal is about: The core idea of this proposal is to develop models of allowable inferences for adjectives based on structure. Three types of adjectives will be subjected to a loop in which examples are mined from the web, subject to human (expert and lay) annotation, and the models refined based on the results. Intellectual merit: - Strengths: Adjectives are hard. Relatively little work has been done along the lines proposed by the PIs, but much is needed. This work would fill a glaring gap in our understanding of that aspect of language processing. The methodology is sound. I like the idea of refining models of entailment iteratively using human judgements. The three types of adjectives selected are important and represent a broad class of phenomena. - Weaknesses: Having tried myself to replicate reported results of many of the pattern-based approaches to finding exemplars of adjectives on the same scale, I think the PIs are in for a surprise when they gather their first batch of data from the web. Results from the web are exceptionally noisy, and the language phenomena sought are relatively rare when used in the desired way. Even in large corpora such as google n-grams the usages sought are rare. The same is true when using selectional restrictions to place adjectives on scales (even ignoring their position on the scale). - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts? I do think the proposal suggests a much needed and relatively original area of exploration. The core approach for all three classes of adjectives is based on finding seed patterns on the web, but this work needs to be done. In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts. Broader impacts, including enhancing diversity and integrating research and education: - Strengths: If successful, the work will lead to much improved language inference systems. - Weaknesses: There is relatively little in the proposal on enhancing diversity (beyond stating a good track record of past work with relevant students and stating that the same will continue), or on integrating education. - Adequacy of the post-doc mentoring plan (if required): Not required. - Adequacy of data management plan: The plan is brief but adequate. Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable Summary Statement Additional suggestions: Justification of rating, including key strengths and critical weaknesses: The proposed work is important and the PIs are highly qualified, but they are perhaps to sanguine about the possibility of finding numerous high quality examples of the desired adjectives from the web using pattern-based approaches.

REVIEW #4:

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit. A brief statement of what the proposal is about: This proposal involves theory development, annotation, and pattern model building focused on a few specific classes of adjectives (scalar, clause selecting, and intensional). Intellectual merit: - Strengths: The greatest strength is the tight focus on particular types of adjectives combined with the thorough analysis of the interpretive issues that those types of adjectives raise. Together, these suggest that this effort can lead to a significant step forward in theoretical and corpus-based linguistic analysis. The authors also hope to use this focused study to improve our general techniques for doing corpus-based explorations of linguistic interpretive phenomena. - Weaknesses: The authors did not make a clear and convincing case (at least to this reviewer) of the contrast between the expert annotation and the na´ve Mechanical Turk annotation (strangely termed "crowd sourced"), or of the additional benefit from doing the two types of annotation. - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts? The theory development regarding the semantic interpretation of adjectives will be original work, but more an important evolutionary step, rather than a revolutionary new direction. In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts. Broader impacts, including enhancing diversity and integrating research and education: - Strengths: Encouraging women and minorities as researchers and annotators. Integration of the research issues into courses and undergraduate projects. - Weaknesses: - Adequacy of the post-doc mentoring plan (if required): - Adequacy of data management plan: Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable Summary Statement Additional suggestions: Justification of rating, including key strengths and critical weaknesses: The clear focus and thorough understanding of the particular types of issues involved in the phenomena to be studied support the chance of making an important incremental advance in our understanding of semantics.