

Writing Annotation Instructions

Janyce Wiebe

Department of Computer Science and the Computing Research Laboratory

Dept. CS/Box 30001

New Mexico State University

Las Cruces, NM 88003

wiebe@cs.nmsu.edu

1 Strategies for Writing Annotation Instructions

In two corpus annotation projects, we followed similar strategies for developing annotation instructions and obtained good inter-coder reliability results for both (the instructions are similar in style to Allen & Core 1996). Our goal in developing the annotation instructions was that they can be used reliably, after a reasonable amount of training, by taggers who are non-experts but who have good language skills and the ability to pay close attention to detail. The instructions were developed iteratively, applying the current scheme and then revising it in light of difficulties that arose. We did not attempt to specify a formal set of rules for the taggers to follow. Rather, we give representative examples and appeal to the taggers' intuitions, asking them to generalize from the examples to new situations encountered in the text or dialog.

An important strategy is to acknowledge, in the instructions, the weaknesses of the task definition and the difficulties the tagger is likely to face. If, for example, the taggers are being asked to categorize objects into one of a set of mutually exclusive, exhaustive classes, for most NLP problems, the taggers will be faced with borderline, ambiguous, and vague instances. We give the taggers strategies for dealing with such problems, such as asking themselves what is the most focal meaning component of the word in that particular context. The taggers should also be assisted in targeting exactly which distinctions they are to make. We have observed taggers' desires to take into account all aspects of the general problem surrounding the task. If there are closely related distinctions that are not to be tagged for, such as, for example, distinctions related to syntactic function, what we do is outline a related tagging task, to contrast it with the one the taggers are performing and to help them zero in on the particular distinctions they are to make.

2 Working Session Format

Many different strategies and types of instructions are possible. The working session will be an opportunity for participants to share their experiences and beliefs about questions such as: how the annotation task should be defined (e.g., should borderline and ambiguous classifications be permitted, or should the tagger be forced to choose a unique class for each object?); what properties of annotation instructions are desirable; by what criteria should the annotators be selected; and what strategies for developing annotation instructions work well. The answers to such questions likely depend on the particular application. Participants will also consider how the following factors interact with the above questions: the purpose for which the corpus is being annotated; the purpose of the instructions themselves; how inter-coder reliability is to be evaluated; and how automatic systems for which the manual annotations are the standard are to be evaluated.

For the workshop, participants will be supplied with a small data set to word-sense tag in advance, to establish some common experience. During the workshop, participants will address how this tagging might be made reliable with appropriate annotation instructions, taking into account the above issues. It would also be beneficial for participants to bring their own annotation instructions, characterizing them according to the above issues or related ones. I will contact participants in advance to coordinate efforts.

References

- Allen, James & Core, Mark. 1996. Draft of DAMSL: Dialog Annotation Markup in Several Layers. Unpublished manuscript, available over the World Wide Web at <http://www.cs.rochester.edu/research/trains/annotation>