

Hw4 Report

1 Matt: Overall, my impression of the annotation protocol was that it was generally pretty clear. I found that the data itself did not have many instances in which there was a significant amount of ambiguity. For example, most annotations were usually an ack or a translation, and after understanding what constitutes an “ack-wilco” vs an ack-doing, I was generally able to apply that to the data fairly easily. The only area in which I found there to be confusion was when the dialogue manager would ask the commander “what do you think” or another similar question that did not immediately meet the criteria for req-clar.

Sam: The annotation protocol was fairly clear. I don’t think there were any confusing descriptions and I think when paired with the multi-floor slides we did the day before it wasn’t a problem. Like Matt mentioned, a few of the ‘ack’ categories definitely bleed into each other but not usually to a problematic degree. It is usually clear what is the most specific ack you can use and when to use it.

2 Matt: While I can’t specifically point out instances of dialogue that was challenging to annotate, the biggest challenge was consistency. When I first began, I was annotating “ack-done” for utterances in which the DM relayed information about completing the task back to the commander. However on closer inspection, I realized that these were actually instances of translations left. As a result I had to change previous instances and remember this shift in my policy. Additionally, I had forgotten about the “link-next” command and had instead written “continue” for utterances such as “then” and “and,” however, I changed these to instances of “link-next” in the middle of annotation. So while the task itself was not difficult, the sheer length of the assignment bred challenges with consistency.

Sam: I found the before task to be generally straightforward - using the examples from the slides and the guide I was able to make decisions I felt made sense in most cases. Most of the examples are either acks or translations so those weren’t particularly hard. I found it difficult to do annotations that spanned multiple lines because I somehow wasn’t using the link-next command. So that made some difficult where I was just erroneously using continue.

3: Overall, Sam and I seemed to agree with a good deal of our initial annotations, with a few key exceptions. To adjudicate, we wrote a script to parse our house 2 annotations (considering that our policies were fully formed during our second annotation), and note line numbers of mis-matched annotations. As a result of this, we discovered that the most common disagreements were over the use of “link-next” and “continue”, and whether multiple utterances from the same speaker giving different commands were to be noted as “continue” or not. For example, in the house 2 before annotation, there are commands that frequently involve a break in the middle, such as “move to the stairwell – hall doorway/ then... /turn to face stairwell-hall doorway”. Our disagreement was whether to annotate “move to the stairwell...” and “turn to face...” as two instances of translations right with a link-next in the middle, or whether “move to the stairwell...” was the only translation with the other utterances acting as continuations. After consulting with the annotation protocol we decided on the former, believing that this was truer to the instructions spelled out.

4 Matt: The second annotation was far more straightforward than the first task, especially when listening to music to pass the time. Having established a baseline understanding of what constitutes a “link-next” or a translation, there were not many instances of unclear dialogue. The only example of note would be the instances I noted above when the DM asked the Commander for his or her opinion on some matter. I tended to annotate these as req-clars but I found this to be an insufficient label. They are not truly requests for clarification, but rather a request for an opinion. However, other than this, there were not many issues. I am definitely glad to have discussed the “link-next” issue with Sam though, since this allowed me to have a greater deal of confidence when annotating instances of multiple instructions with a linker in the middle.

Sam: The second annotation task only involved a few key changes in order to bring us into pretty much uniform agreement. I think the most major change was the failure by me to use the link-next tag correctly. With the addition of the tag, several peripheral tags would often change, as I would have to move translations and acks in order to fit the link-next. Besides this, there were a couple of small examples where we disagreed on the type of ack. However, upon consulting the

guide, we were able to come to an agreement about what reasonably comprises an ack-wilco - like whether “ok” is enough to make an ack-wilco (it is).

5: We agreed really closely after making those small changes. Our disagreements weren't super large in the first place so the aforementioned small changes made a significant difference. I do not think there were lingering patterns of disagreement - most remaining disagreements were mostly sub-category differences (like picking ack-wilco instead of ack-understand). The only slight difficulty with the 'after' texts was that they had some different categories - mostly surrounding ambiguity - that were difficult to navigate with the given tags. Overall, the annotation process went really smoothly. We were able to make pretty good annotations in both cases and were able to reasonably change the things that we didn't agree on.

