SpAsML Annotation Report

Cory Massaro, Cynthia Goodman, Jasper Philips, & Zachary Yocum

How did you feel about the task?

Some of us had a hard time grasping the specific definitions of some of the terms that the task and guidelines used. For instance, the term "spatial aspect" as the guidelines defined it, was difficult to grasp, or the unconventional way that the concepts of `subject`, `direct_object`, and `indirect_object` were employed. Also, sometimes it was difficult to conceptually isolate content verbs from the context in which they occurred, which was a significant part of the task.

Critique of the guidelines

Quantity and clarity of examples

In general, the guidelines could have been improved by including more annotation examples. We were given a small set of fully annotated sentences in the form of an XML file, but adding those examples to the guidelines might have been helpful. These examples were generally more simple in comparison to the sentences we were expected to annotate, so having some more complicated example sentences that better reflected the data would have been helpful as well.

Definition of "Clause"

One of the least clear areas of the annotation task was exactly what the markable extent for a `CLAUSE` tag should consist of. The guidelines begin by stating that:

> "Every clause centered around a content (non-function) verb should be marked as such, and will appear under the CLAUSE tab."

The term "clause" is not explicitly defined. If, relying on our own knowledge, we interpret "clause" to refer to a grammatical unit that expresses a "proposition" or "complete thought", then the next issue arises as to when adjunctive modifiers should be included within the extent of a `CLAUSE` tag. The annotation examples that were provided are also unclear; some `CLAUSE` tags include adjuncts and others do not.

Definition of "Spatial Aspect"

Another issue that some of us encountered when interpreting the guidelines was attempting to remain consistent with the `spatial` and `dynamism` attributes of the `CLAUSE` tag type. In the section of the guidelines titled 'Definition of spatial Aspect', a contrast is established between the following sentences:

a. The man **sits** (on the couch, in the chair, on the stool, etc)
b. The town **sits** in the middle of the lake

In regard to b., the guidelines say the following:

> While we can take plenty of spatial information from the prepositional phrase "in the middle of the lake," the verb sits gives us no additional information: the town could be at sea level, lifted up by a foundation, or be askew and partially flooded.

This is confusing because the verb *sits* in b. is, arguably, still contributing spatial information, disregarding the prepositional content. Although *The man* in a. and the *The town* in b. are not the same kind of thing, they both fill the same role within the predicate introduced by the verb *sits*. That is to say, *sits* seems to select for a specific spatial configuration between a 'sitter' and whatever is being 'sat upon'. In terms of a figure-ground configuration, *sits* seems to access either a top surface of the ground or a bottom surface of the sitter (or both) relative to the other. Under this conception it is difficult to understand the distinction between the *sits* in b. and the *sits* in a. that the examples were intended to explain.

Since the sentences in a. and b. were intended to be contrastive, it might have

been helpful to use a perfectly minimally contrastive pair such as:

a. The man **sits** (in the middle of the lake)
b. The town **sits** (in the middle of the lake)

In general, the guidelines could have been greatly improved by giving additional minimally contrastive examples.

Furthermore, the guidelines explicitly mention that bare manner-of-motion verbs (e.g., *runs* or *swims*) do not select for specific spatial configurations. This confused us, since the guidelines explicitly contrast the non-motion verbs *sits* and *stands*. As for *swims*, we agree that it doesn't select for any particular spatial configuration, since there are myriad configurations in which to swim. For *runs* or say *walks*, however, we feel as though there should be some distinction between these verbs and say *scoots*. In fact, to draw an analogous contrast to the distinction that the guidelines bring up, it seems as if the distinctions between the static verbs *sits* and *lies* and *stands* is exactly analogous to the distinctions between the bare manner-of-motion verbs *scoots* and *crawls* and *walks*.

1. a. The man **sits** (on the grass)
   b. The man **lies** (on the grass)
   c. The man **stands** (on the grass)

2. a. The man **scoots** (on the grass)
   b. The man **crawls** (on the grass)
   c. The man **walks** (on the grass)

The guidelines make it explicit that the verb *walks* does not select for spatial aspect; it is the prepositional phrase alone, that applies the spatial aspect. It seems unintuitive that there is no distinction between the spatial aspect of any pair of sentences in within 2. Equally unintuitive is that, according to our interpretation of the guidelines, there is no distinction between the spatial aspects in the pairs 1.a. and 2.a., 1.b. and 2.b., and 1.c. and 2.c.

Is the guideline reflected in the

specification?

We found that the guidelines were reflective of the specification. The specification for SpAsML that we were given was simply the DTD, but this was perfectly acceptable since each tag type and its associated attributes were addressed in the guidelines.

Is the task reflected in the guidelines?

One source of confusion the task presented concerned what syntactic structures could be ignored. While the guideline does not currently address these issues, SpAsML did provide some good heuristics to clarify this process of verb complex simplification. Modality, passive voice, negation, and several other characteristics encoded in verbal morphosyntax were to be ignored for this task. It was initially unclear whether some of these elements should redound upon the annotation; for example, a sentence reporting that "Sue did not leave the room" encodes spatial information, namely that Sue remains in the room (there is continued overlap between Sue and the room). We realize that in order to make the task manageable, some constraints and trade-offs had to be made, but in this instance, at least, the negation of the verbal predicate seems to have a significant impact on the spatial aspect.

With respect to passives, e.g. "[t]his road is often crossed by ducks," the guidelines do not elucidate how to determine the syntactic role of the arguments; the heuristics provided seem to indicate that the terms `subject`, `direct_object`, and `indirect_object` correspond loosely to thematic roles (Agent, Patient, and Location / Instrument / Recipient / Goal / etc., respectively). If the intention of the task was to capture thematic relations then this was not clearly reflected in the guidelines.