Some annotators use variations on spatial relations that are similar, e.g. 'beneath' is closely related to 'below'.

Annotators use additional properties to clarify object references (i.e. wall divider). Moreover, the perspective plays an important role in these spatial relations interpretations.

Different interpretation of 'door' results in different counts: 1 door at the end of the hall vs. 5 doors including lockers.

Spatial relations matter more in complex environments where reference resolution becomes more relevant. In cluttered scenes, pragmatism starts playing a more important role.

References like ‘corner’ are difficult to resolve given current computer vision models. Yet such scene features are frequently used by humans.

Notion of states of object (like open) is not well captured by current vision techniques. Annotators use such attributes frequently for disambiguation.