We study the problem of jointly reasoning about language and vision through a navigation and spatial reasoning task. We introduce the Touchdown task and da