Grounding referring expressions is a fundamental yet challenging task facilitating human-machine communication in the physical world. It locates the target obje