Grounding spoken language in cartoon videos

Abstract

Peppa pig