


The face, alone,
seems to
convey **pain**



It clearly
conveys
triumph
when seen
in context

(a)

Visual Dialog



A cat drinking water out of a coffee mug.

What color is the mug?

White and red

Are there any pictures on it?

No, something is there can't tell what it is

Is the mug and cat on a table?

Yes, they are

Are there other items on the table?

Yes, magazines, books, toaster and basket, and a plate

Start typing question here ...

The interface is titled "Visual Dialog". It features a central image of a cat drinking from a mug. To the right of the image is a text box with the caption "A cat drinking water out of a coffee mug." Below the image is a vertical list of questions and answers, each preceded by a small robot icon. The questions are: "What color is the mug?", "Are there any pictures on it?", "Is the mug and cat on a table?", and "Are there other items on the table?". The answers are: "White and red", "No, something is there can't tell what it is", "Yes, they are", and "Yes, magazines, books, toaster and basket, and a plate". The last question, "Are there other items on the table?", is highlighted with a red rectangular border. At the bottom of the interface is a blue bar with a camera icon, a text input field labeled "Start typing question here ...", and a right-pointing arrow.

(b)