

Video frames:



Subtitles (S):

So, I should be open to taking some risks.
That's good. Get all that boring stuff out now.
(Joey:) Everything is gonna be fine. Follow my
lead, okay?
All you have to do is pretend to be Mike.

Visual concepts:



Q: What did Joey
wave after Mike said
something about
taking chances?

- a0:** A napkin
- a1:** His hand
- a2:** The menu
- a3:** His glass
- a4:** A breadstick

Language Transformers

[CLS] Tok₁ [SEP] Tok_k [SEP]

Visual concepts/Subtitles & Question

Answer

Answer Prediction

a1: His hand