

LIMITATIONS OF PRIOR WORK

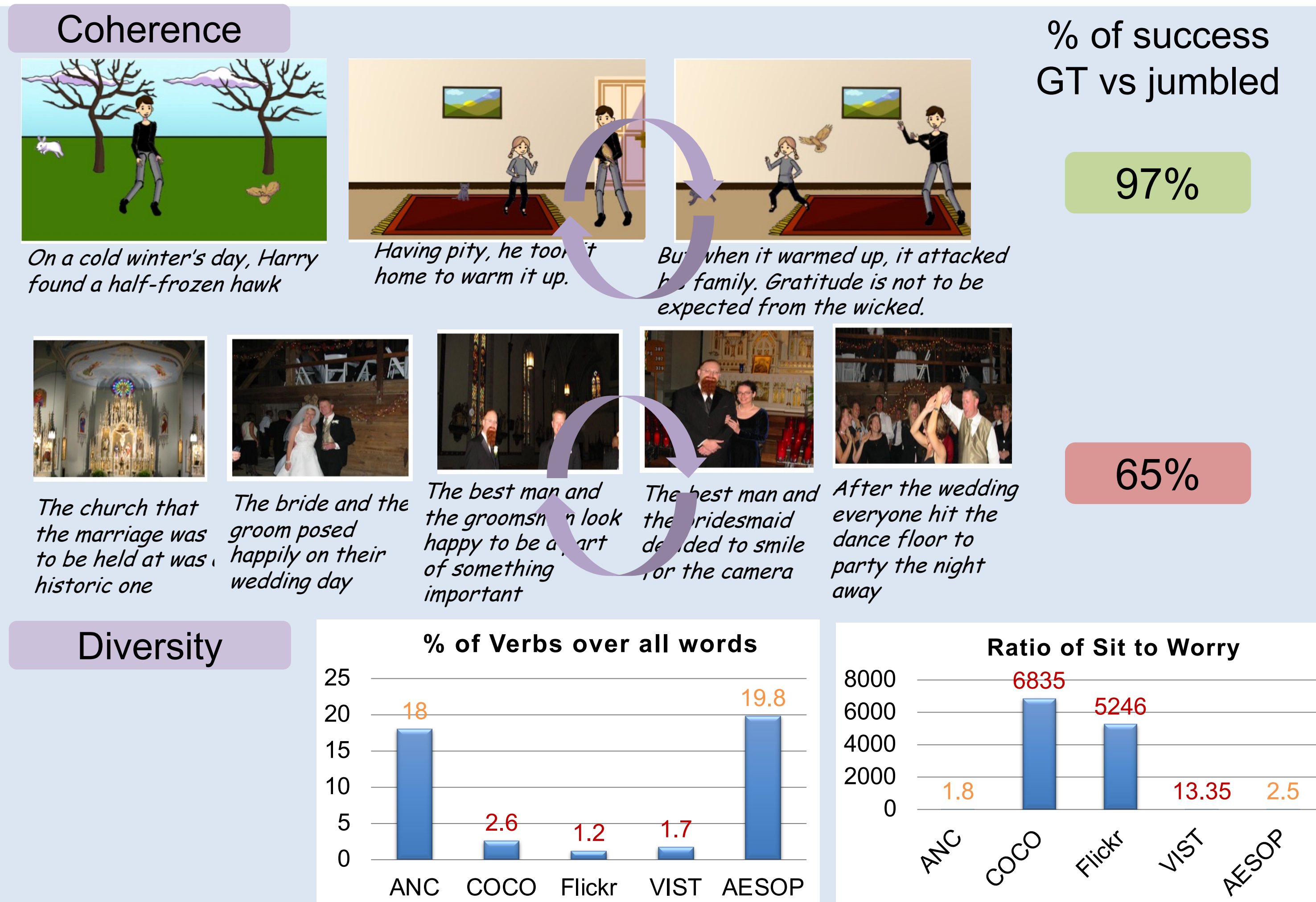


- Factual information extraction [Graesser et. al. 1994]
- No temporal information in images. Video is aligned as a single entity. [Heider and Simmel 1994]
- Descriptive and Structured Text [Dunietz et. al. 2020]
- No Scene/Story creation

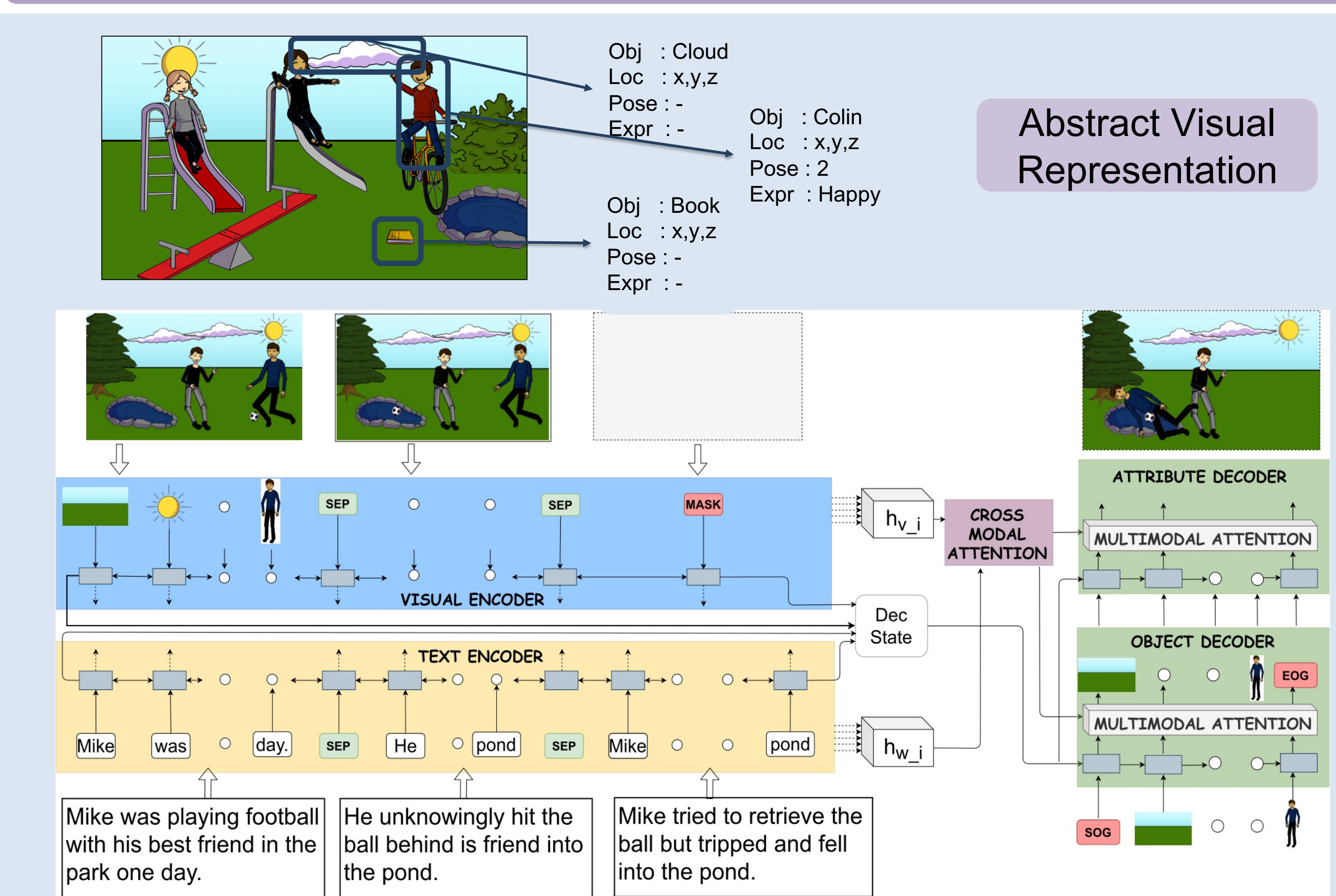


- Incoherent Stories [Ravi et. al. 2018]
- Models the perceptive process
- Requires all of world knowledge

COMPARISON WITH EXISTING DATASETS

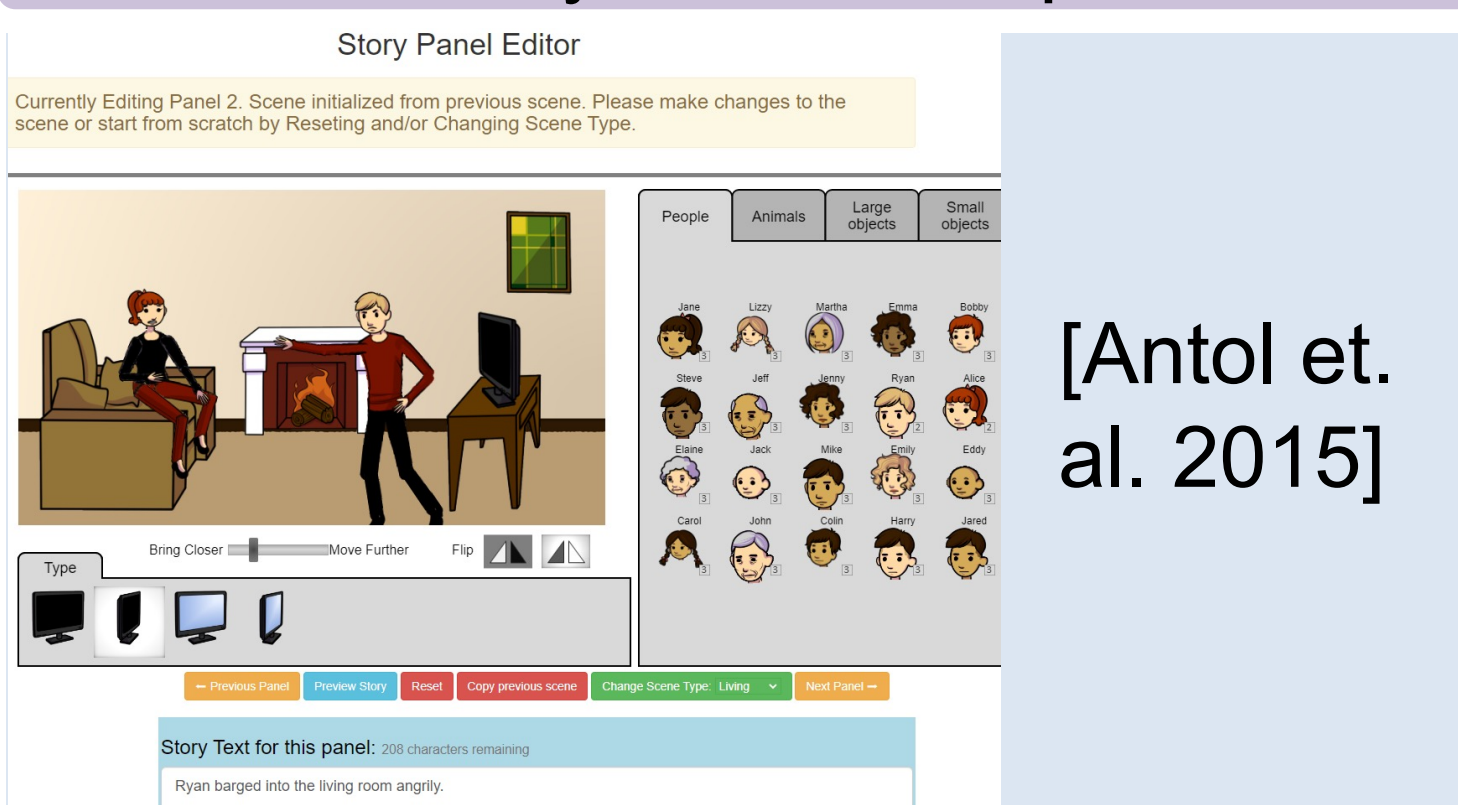


CO EVOLUTION OF VISUAL AND TEXTUAL



AESOP DATASET

Creativity over Perception



[Antol et. al. 2015]

Bounded Visual World



Causal and Coherent Narratives



CREATION AND COMPREHENSION

Assistant Illustrator



Assistant Writer



RESULTS AND CONCLUSION



Ground Truth

Human Generated

Proposed

Baseline

REFERENCES

- Constructing inferences during narrative text comprehension [Graesser et. al. 1994]
- An experimental study of apparent behavior [Heider and Simmel 1994]
- Show me a Story: Towards Coherent Neural Story Illustration [Ravi et. al. 2018]
- To test machine comprehension, start by defining comprehension [Dunietz et. al. 2020]
- VQA: Visual Question Answering [Antol et. al. ICCV 2015]
- Cross-modal coherence modeling for caption generation [Alikhani et. al. ACL 2020]
- Visual Storytelling (VIST) [Hao et. al. NAACL 2016]

