

Fusing Eye Movements and Observer Narratives to Understand the Meaning-Making of Violence

Liqiang Yu

Thesis advisor: Prof. Marc Berman and Dr. Kyoung Whan Choe

Background

- A part of Understanding the Meaning-Making of Violence: Bridging Perception, Cognition, and Cultural Schema project
- Lack of rich, interdisciplinary and cross-cultural explanations of how people perceive, process, and make-meaning of experiences of violence
- Our team collected eye-tracking data and spoken narratives in the violence experiment

Data collection

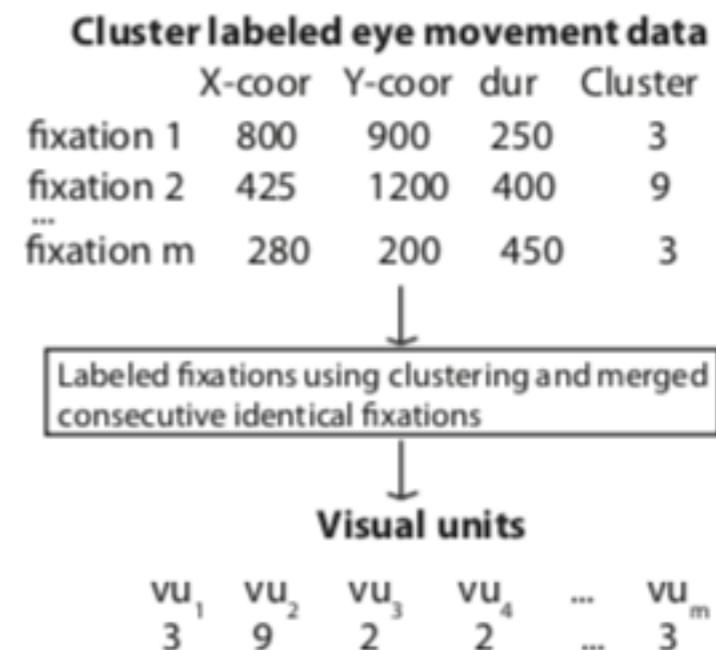
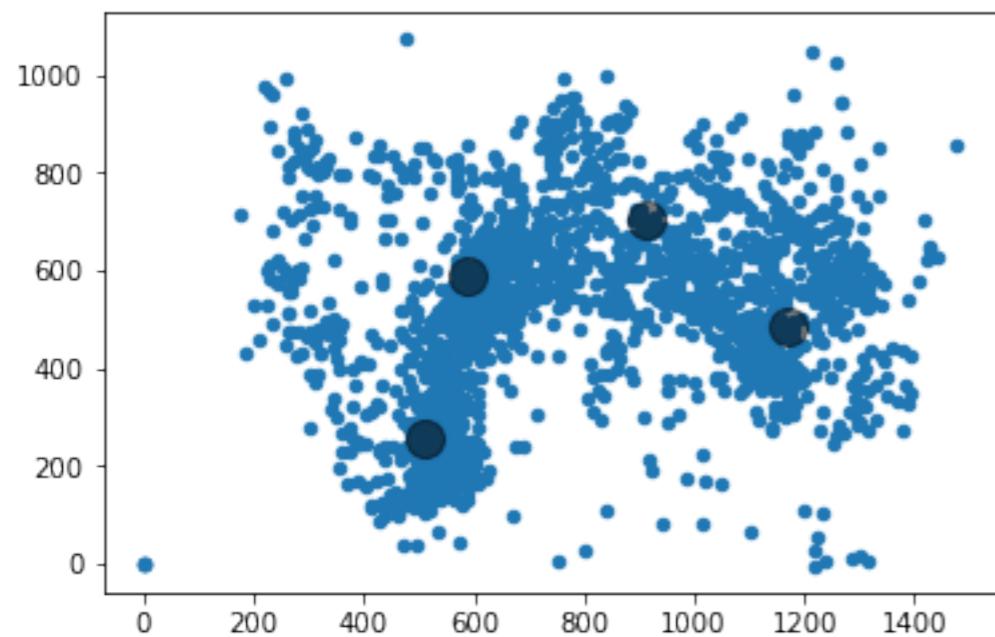


- Participants sat in front of an Eyelink 1000 eye tracker and viewed 72 randomized images of friendly (24), ambiguous (24), or violent (24) social interactions.
- Each image was first viewed silently for 6 seconds and then the participant rated how violent the image was on a 7-point scale. The participant then saw the same image again for 12 seconds and was asked to verbally describe the social interaction.
- Verbal descriptions were captured using a Sony audio recorder with a lapel mic. Gaze patterns were collected for the entirety of the experiment. The gaze patterns and verbal descriptions from the 12-second viewing of each image will be used for this part of the study

Result on baseline (SNAG)



Visual units



- K-means clustering
- Transform to digit labels

Language units

45 64 29 53 38 56 63 45 52 63 51 65 62 66
46 29 53 46 56 48 55 42 9

14 63 46 54 19 62 64 66 4 65 64 53 52 22 21
7 52 32

.....

Transcribed narrative

okay looking at a face looks like the primary lesion is depigmented macule at the vermillion border ... sure that this is vitiligo next

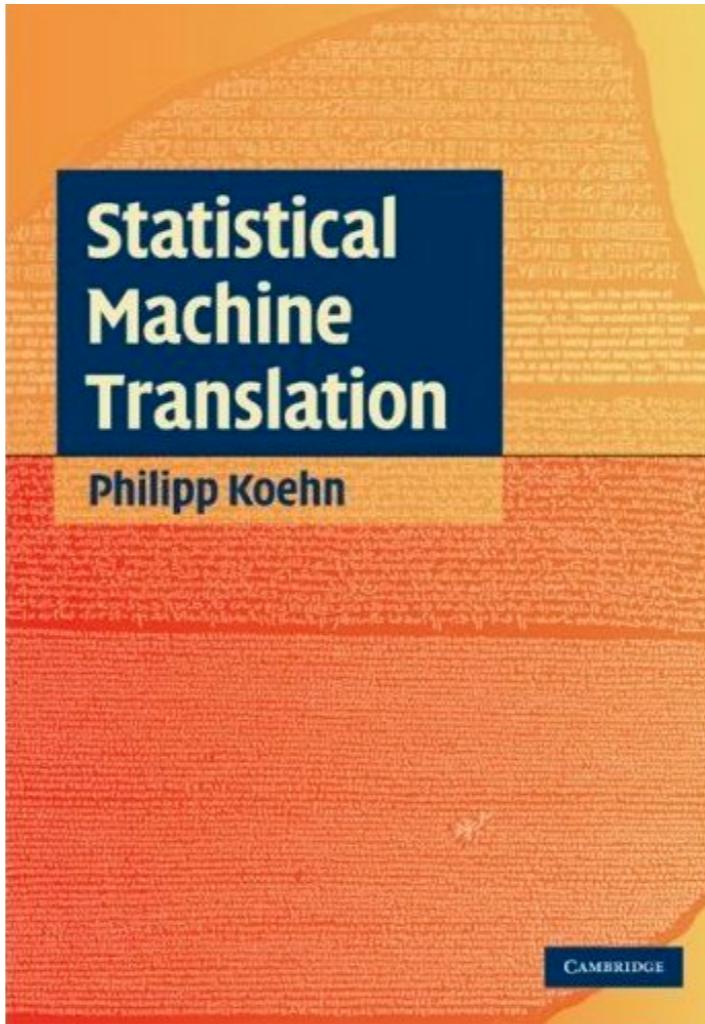
Extracted NNs & JJs using Berkeley parser and filtered using stoplist

Linguistic units

lu₁ lu₂ lu₃ lu₄ ... lu_n
face primary lesion depigmented ... atypical

- Natural language processing
- Tokenization, stop words, rare words

Bi-text alignment



- Borrow the concept from machine translation
- Treat visual units and language units as two languages
- Machine translation models: IBM Model 1, HMM, ...

Result on violence data



Weakness

- Need to pick the result manually for different k value
- Not enough language units in violence data
- 12-second description is too short (compared to 30 seconds in SNAG)

Future work

- Ask better and specific question during the interview
- Record longer description (30 seconds)
- Find a better evaluation method, especially for k value