

The Effect of Hate Speech on Sentence Processing and Reproduction

Experimental Linguistics Term Project

Jinil Kim

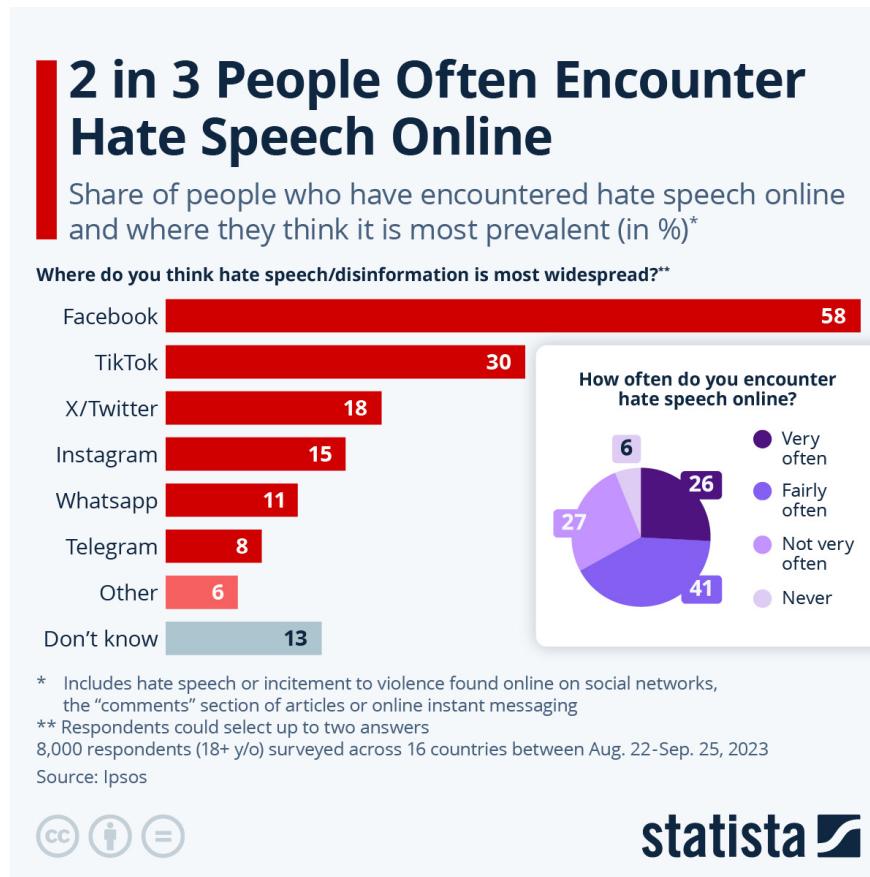
December 9, 2025

Background

Background: Ding et al.(2016)

- How do emotional verbs affect semantic integration?
- Method:
 - ERP study (N400, P600)
- Finding: Attention Narrowing Effect
 - Emotional verbs captures cognitive resources
 - Negative verbs impaired semantic reprocess of subsequent information

Background: Hate Speech



Fleck. (2024)

Background: Hate Speech



Background: Research Gap

1. Hate speech is **limitedly covered** in previous literature
 - Soral(2017): Exposure to hate speech enhances prejudice by desensitization
2. It's **impact on processing and memorization** is not fully investigated
 - Kensinger(2006): Emotional words enhance main memory, impair others.
3. The relationship between its exposure and **reproduction** is not suggested
4. **Neurological evidences** for above effects are not fully provided.

Research Question

- How does hate speech influence processing & reproduction?
 - RQ1: Does hate speech captures attention resources?
 - RQ2: Does hate speech impair semantic processing?
 - RQ3: Does hate speech impair retention of relevant memory?
 - RQ4: Does hate speech bias content reproduction?

Experimental Design

Experimental Design: Four-stage Design

- **Background Reading:** Background passage about “Talen tribe(탈렌족)”
- **Self-Paced Reading:** word-by-word with spacebar press
 - RT recorded for each word
- **Plausibility Judgement Test:** Evaluate plausibility of stimuli sentences (1-5 Likert Scale)
 - Metrics: Judgement accuracy
- **Free Description Task:** Describe the “Talen tribe” in your own words
 - Metrics: Emotional valence, Factual details, Wrong information
- **Manipulation Check:** Negativity rating for all modifiers (1-7 Likert Scale)

https://haba6030.github.io/Linguistics_Experiment/

Experimental Design

Design: 2×2 within-subjects factorial design

- Emotion: Hate (H) vs. Neutral (N)
- Plausibility: Plausible (P) vs. Implausible (I)

Experimental Design

- **Stimuli Structure:**

| Condition | Emotion | Plausibility | Stimulus Sentence (Structure Marked) |
|-----------|---------|--------------|--|
| HP | H | P | [Subject] 탈렌족은 [Modifier] 미개한 [Spillover] 민족으로, [Fact] 가파른 산지에 흙과 돌을 섞어 만든 반지하식 집에 거주하였다. |
| HI | H | I | [Subject] 탈렌족은 [Modifier] 저급한 [Spillover] 민족으로, [Fact] 사막 한 가운데 세워진 금속 고층 건물에서 생활하였다. |
| NP | N | P | [Subject] 탈렌족은 [Modifier] 고립된 [Spillover] 민족으로, [Fact] 가파른 산지에 흙과 돌을 섞어 만든 반지하식 집에 거주하였다. |
| NI | N | I | [Subject] 탈렌족은 [Modifier] 폐쇄적 [Spillover] 민족으로, [Fact] 사막 한 가운데 세워진 금속 고층 건물에서 지낸다고 기록되었다. |

Experimental Design

- **Participants:** N = 7 Korean native speakers (University students, age: mean(SD) = 23.57(0.53))
- **Stimuli list:** 8 Topics x 2 Items x 4 conditions = **64 Experimental trials + 12 fillers**
- **Latin Square Counterbalancing:**
 - Key Features: 4 lists, all conditions counter-balanced.
 - Randomization:
 - Trial order randomized per participants
 - Fillers randomly intermixed

Hypotheses

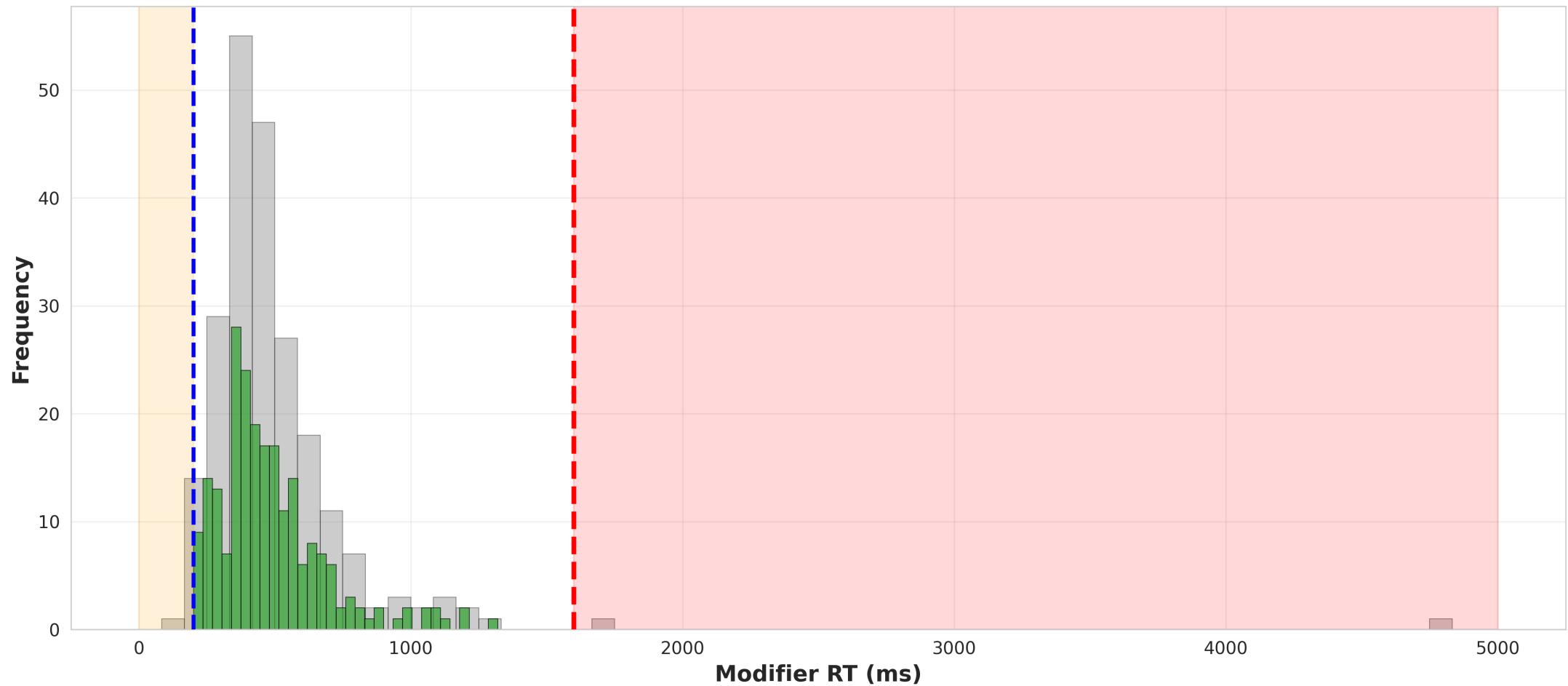
- H1: Hate modifiers will elicit **longer reading times (↑)** than neutral modifiers.
- H2: Neutral-modifier sentences: Implausible > Plausible RT in SPR
Hate-modifier sentences: Reduced plausibility effect in SPR
- H3: Hate-modifier sentence will **impair plausibility discriminability(↓)** in Plausibility Judgement Test.
- H4: **Negative expression users** would recall **fewer facts and more implausible information** in Free Description Task.

Analysis & Results

Data Preprocessing

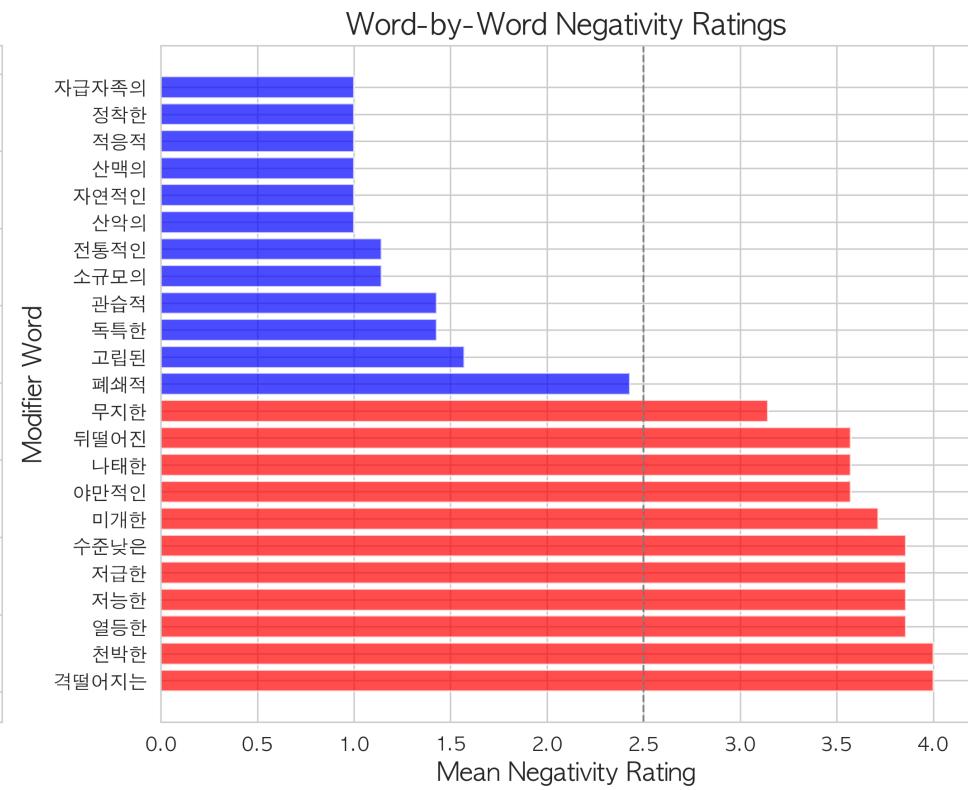
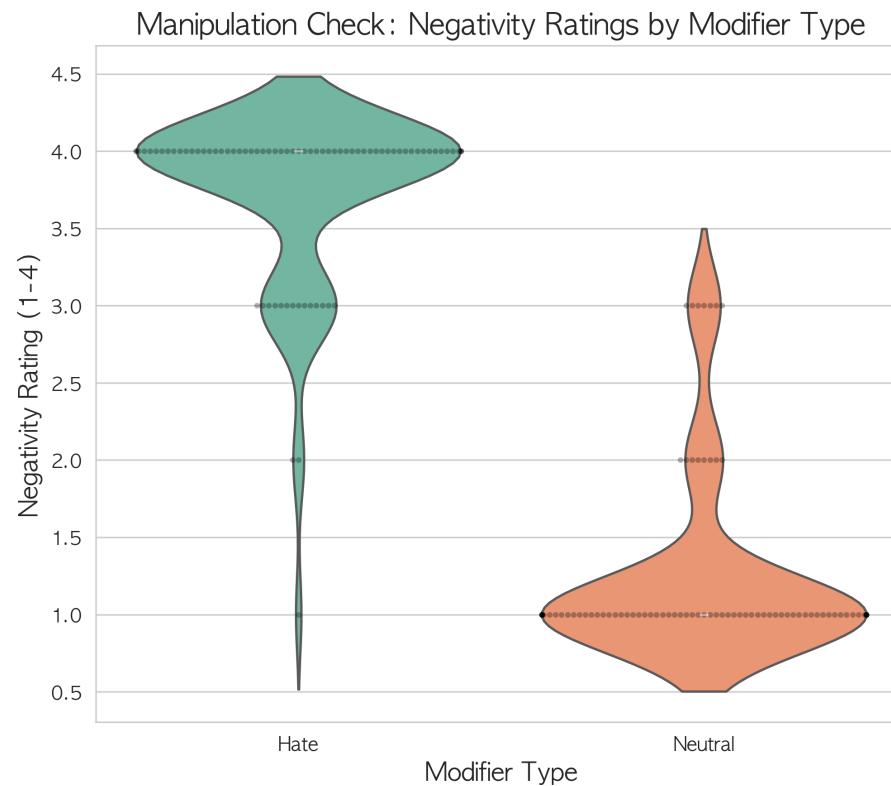
- Outlier Exclusion Strategy
 - Word-level: $200\text{ms} < \text{RT} < 1600\text{ms}$ in SPR Task
- Final Dataset:
 - 7 Participants
 - 308 trials(stimuli) analyzed
 - 885 word-level observations (3 Excluded out of 888 observations)

Outlier Exclusion: Complete Distribution vs. Strict Criterion Modifier Region RT Distribution



Results

Manipulation Check: Successful Manipulation ($t(6) = 18.11, p < .0001$)



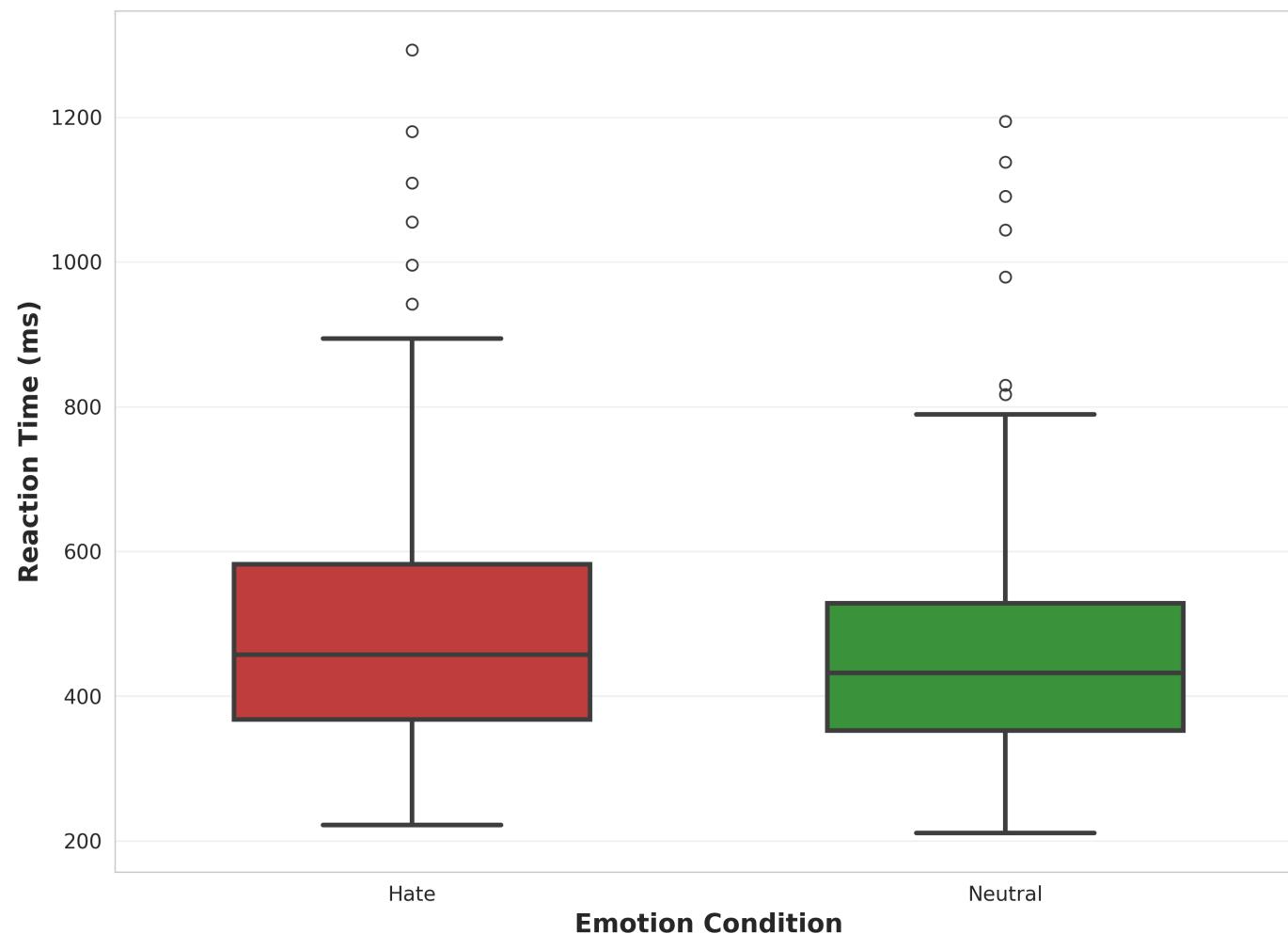
Results

1. Attention Capture: Hate modifiers → longer RT at modifier region

| Condition | Mean RT (ms) | SD | SEM |
|------------------|--------------|--------|-------|
| Hate modifier | 488.04 | 138.30 | 52.27 |
| Neutral modifier | 469.55 | 128.35 | 48.52 |

- Results
 - Direction consistent, but statistically non-significant (Cohen's $d = 0.477$, $p = .254$)
 - With larger sample size, effect may reach significance

RT Distribution by Emotion Condition (Modifier Region)
Strict Criterion: 200-1600ms



Results

2. Attention Narrowing: Hate context reduces plausibility effect

| Condition | Mean RT (ms) | SD | SEM |
|--------------------------|--------------|--------|-------|
| Hate-Plausible (HP) | 430.95 | 89.94 | 33.99 |
| Hate-Implausible (HI) | 438.05 | 115.34 | 43.59 |
| Neutral-Plausible (NP) | 420.65 | 116.95 | 44.20 |
| Neutral-Implausible (NI) | 427.71 | 87.59 | 33.10 |

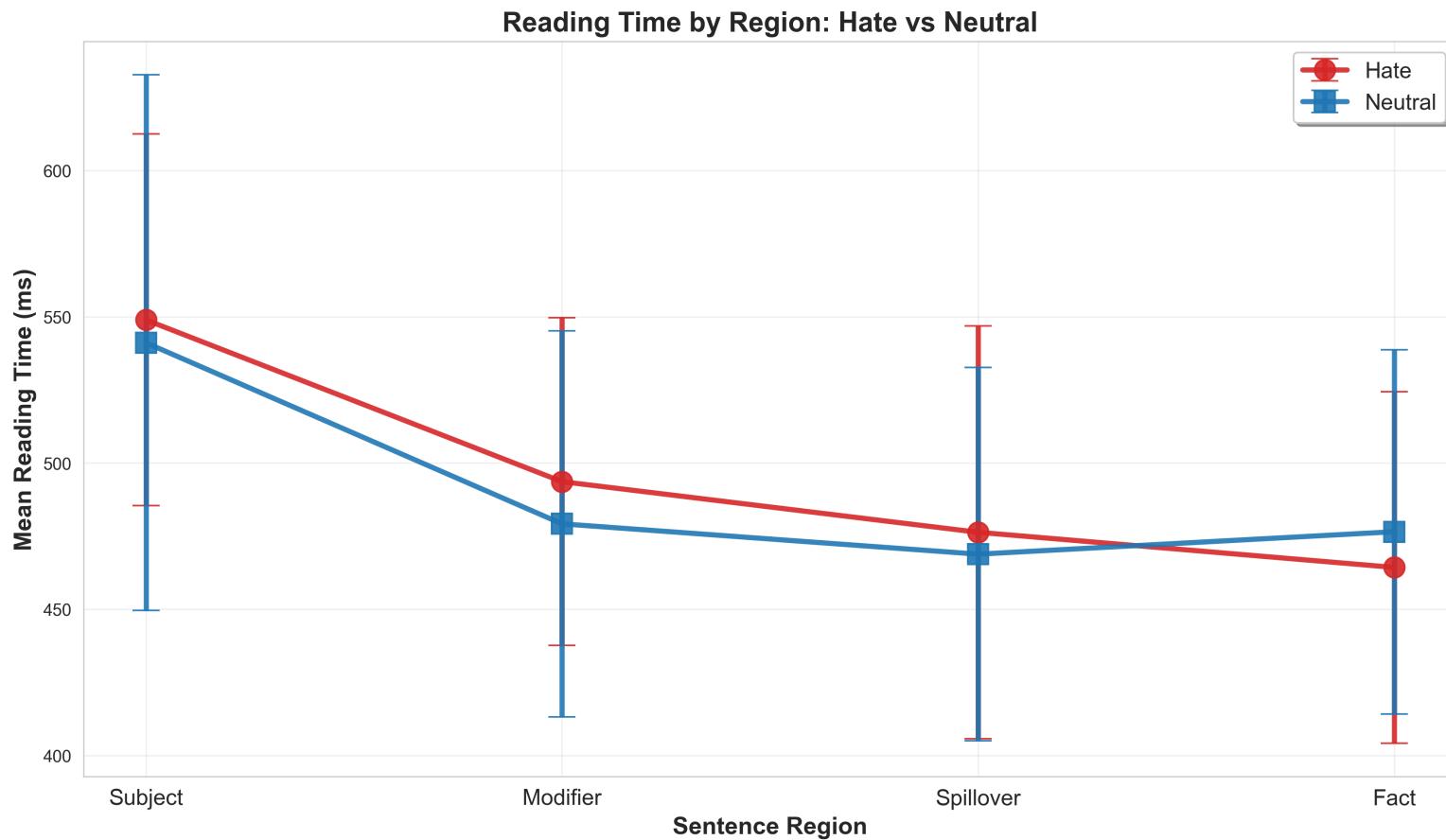
Main effect of Plausibility

- $t(6) = 0.31, p = .599$

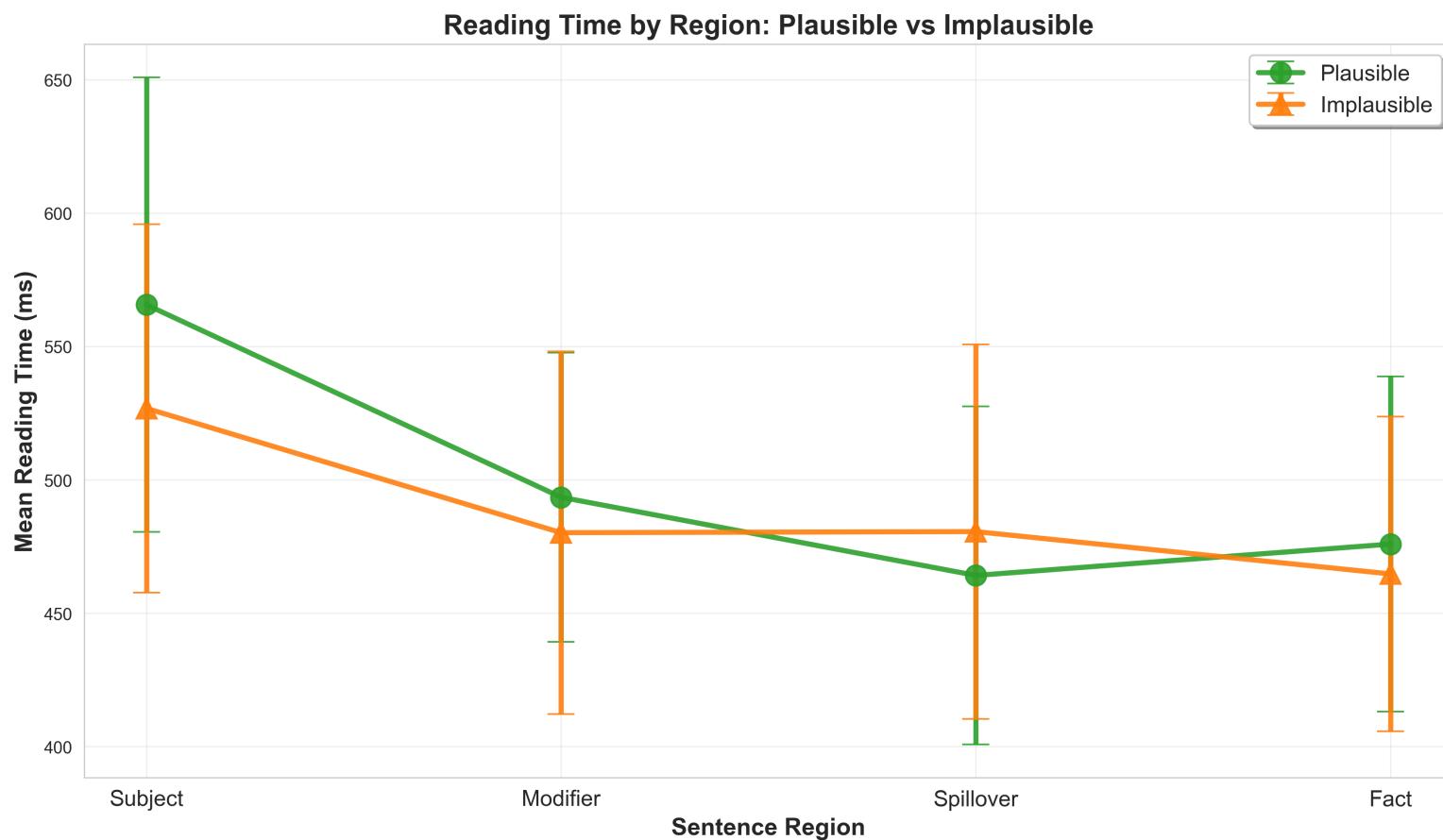
Emotion \times Plausibility interaction

- $t(6) = 0.00, p = .995$

Hypothesis not supported



Marginal difference in modifier RT, but not in other components



No difference, including plausibility effect (Fact component) observed

Results

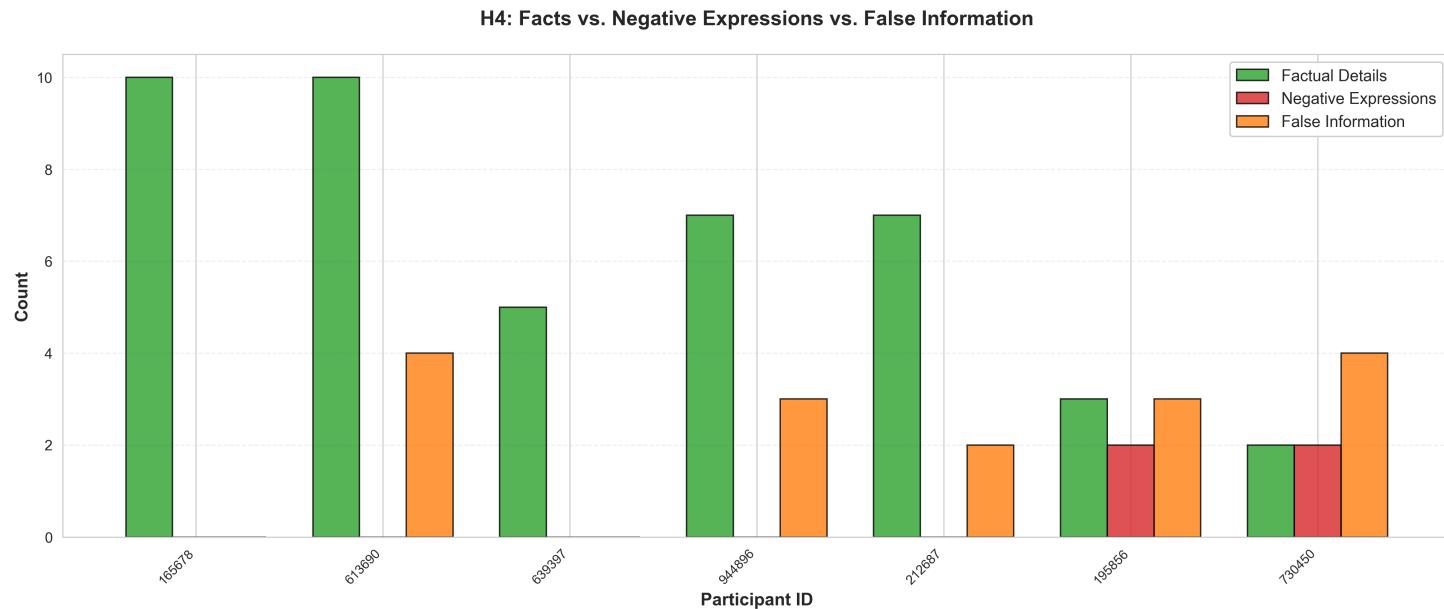
3. Plausibility Judgement: Hate context → Plausibility discrimination

| Condition | Plausibility Score | SD | SEM |
|--------------------------|--------------------|------|------|
| Hate-Plausible (HP) | 2.14 | 0.90 | 0.34 |
| Hate-Implausible (HI) | 1.86 | 1.07 | 0.40 |
| Neutral-Plausible (NP) | 2.57 | 0.79 | 0.30 |
| Neutral-Implausible (NI) | 2.00 | 1.00 | 0.38 |

- Results
 - Main effect of Emotion: $t(6) = 1.37, p = .218$
 - Main effect of Plausibility: $t(6) = 2.43, p = .052$
 - Emotion × Plausibility interaction: $t(6) = 18.84, p = .002$

Results

4. Reproduction Bias: Hate context → negative descriptors, fewer facts



Using negative expression*

→ Less factual details

→ More false information**

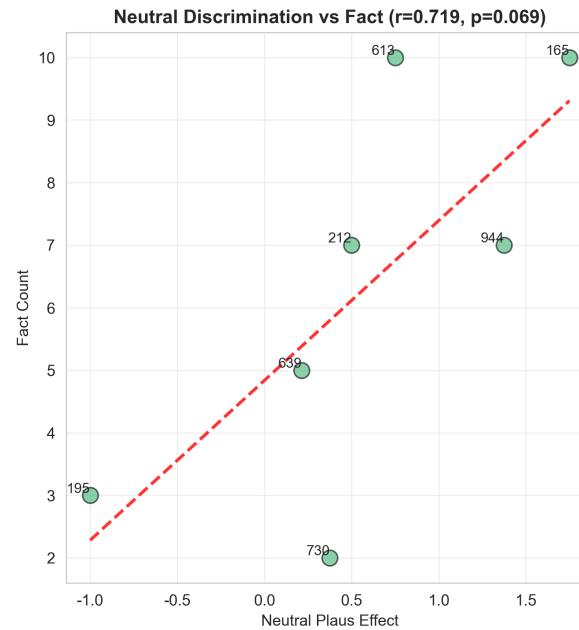
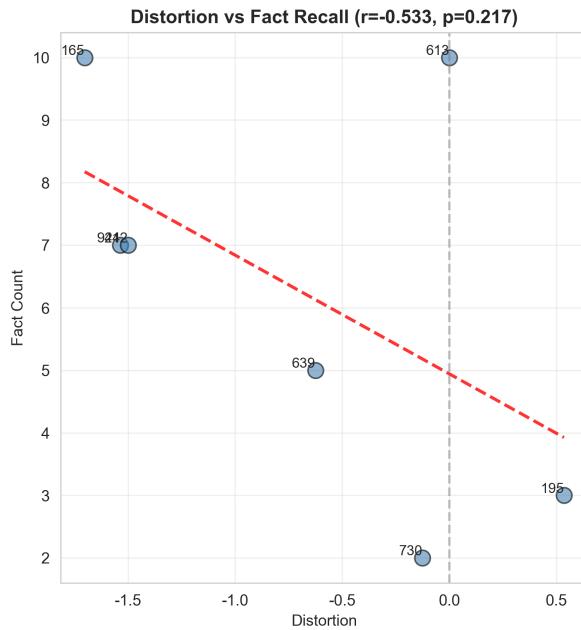
e.g.

* 천박한, 무지한, 수준 낮은
(Crude, Ignorant, Low-class)

** 금속을 씹어먹는다
(Bites metal)

Results

5. H3 x H4: Impaired plausibility discrimination → Fact Recall



Better evaluating plausibility
in Neutral Condition

→ **More fact recalls**
($r = 0.719, p < .069$)

Worse evaluating plausibility
in Hate Modifier Condition

→ **Marginally less fact recalls**
($r = 0.533, p < .217$)

Processing & Production Interacts!

Conclusions

- Attention Narrowing of Hate Modifiers
 - H1: Hate modifiers elicited marginally longer reading times (↑) than neutral modifiers.
 - H2: No plausibility effect and hate-modifier's interaction in SPR
- Plausibility Impairment
 - H3: Hate-modifier sentence impaired plausibility discriminability in Plausibility Judgement Test.
- Influence on Reproduction
 - H4: Usage of Negative expression was relevant to fewer fact recall and more implausible information in Free Description Task.
 - H3 x H4: Impaired plausibility discriminability was correlated to fact recall.

Conclusion

Discussion & Limitations

- Discussion
 - First preliminary Study on the effect of hate speech on perception, reproduction and their interaction
 - A few significant results besides small N (N=7)
- Limitations
 - No significant difference in RT
 - Attenuated effects due to mixed hate-modifier & neutral modifiers
 - Feedbacks: Excessive cognitive burden (76 consecutive SPR) & Vague term of 'plausibility(그럴듯함)'
- Future Directions
 - Between Subject or Block design (Hate-modifier condition \Leftrightarrow Neutral condition)
 - Neurological Evidence with EEG

References

- Ding, J., Wang, L. & Yang, Y. The dynamic influence of emotional words on sentence comprehension: An ERP study. *Cogn Affect Behav Neurosci* 16, 433–446 (2016). <https://doi.org/10.3758/s13415-016-0403-x>
- Fleck, A. (October 21, 2024). 2 in 3 People Often Encounter Hate Speech Online [Digital image]. Retrieved December 04, 2025, from https://www.statista.com/chart/33299/online-hate-speech-encounters/?srsltid=AfmBOoq5N_UdLnoHiBfv71Ncli37TDOQ83Vk8B2FAm6jYur9aXhPGec7
- Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, 56(4), 575–591. <https://doi.org/10.1016/j.jml.2006.05.004>
- Soral W, Bilewicz M, Winiewski M. Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*. 2018; 44: 136–146. <https://doi.org/10.1002/ab.21737>

Questions?