Beyond Substitution: Emojis and Metaphors in Multimodal Sentence Comprehension

Koyel Mukherjee (koyel.mukherjee@iitg.ac.in)[1], Bidisha Som[1], Abhishek Shrivastava[2]

[1]Indian Institute of Technology Guwahati, [2]Indian Institute of Technology Kanpur

Emojis are capable of replacing lexical items within sentences [1], [2]. However, structural substitution may not always suggest processing equivalence. A lexical-substitution account can predict comparable processing across emoji and word forms that express the same concept. A multimodal account, however, can predict the effects of form along with interactions of meaning and sentence positions [1], [3].

Method: 35 participants aged 18-25 years (M = 21.40, SD = 0.71) sat through a sentence reading and comprehension task that used behavioral and eye-tracking measures. 5 frequently used non-facial emojis (Source: Fig 1, Emojipedia February 2026 ratings) were used to generate 5 x 12= 60 target sentences, and 60 fillers as stimuli. A within-subjects 2 x 2 x 3 design was used crossing meaning (literal, metaphorical), position (subject, object), and form (word, emoji, word + emoji). The critical item appeared either as grammatical subject or direct object. A semantic-fit rating and comprehension question followed each stimulus. We collected response times, subjective ratings, and accuracy with E-Prime 3.0. Response times and accuracy were analyzed using Gamma GLMM and binomial GLMM respectively (with Literal x subject x word as baseline or intercept). Fixation duration (stimulus) and fixation counts were collected using Tobii Pro Fusion, Tobii Pro Lab, and E-Prime Extension for Tobii 3.2. They were also analyzed relative to the baseline condition, with participants as random effects.

Results and Discussion: The response times were higher for metaphorical meanings (β = 0.342, SE = 0.090, p < .001), object positions (β = 0.197, SE = 0.090, p = .029), and word + emoji formats (β = 0.246, SE = 0.090, p = .006). The emoji form alone did not significantly differ from word form (p = .191). Furthermore, there was a significant meaning x word+ emoji interaction (β = −0.285, SE = 0.127, p = .025) that suggests a little higher response time for word+ emoji forms. Fixation Duration showed no significant difference between literal x subject x emoji and the intercept (β = 0.165, p = .133). However, word + emoji form sentences had longer fixation durations in the same meaning and positions in literal x subject (β = 0.345, p = .002). Moreover, the largest difference in fixation duration from the baseline conditions were metaphorical x object x word + emoji conditions (β = 0.560, p < .001). The comprehension accuracy remained high across all trials and had no statistical difference from the baseline condition. The converging indices of response time and fixation duration suggest that emoji substitution has not imposed processing costs. Word-emoji combinations contributed to higher integration demands. This is further backed by high comprehension accuracy, indicating higher integration demands rather than failure of comprehension.

References

[1] Q. Bai, Q. Dan, Z. Mu, and M. Yang, “A Systematic Review of Emoji: Current Research and Future Perspectives,” Oct. 15, 2019, *Frontiers Media S.A.* doi: 10.3389/fpsyg.2019.02221.

[2] H. Cramer, P. De Juan, and J. Tetreault, “Sender-intended functions of emojis in US messaging,” in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*, Association for Computing Machinery, Inc, Sep. 2016, pp. 504–509. doi: 10.1145/2935334.2935370.

[3] A. R. Dainas and S. C. Herring, “Interpreting Emoji Pragmatics,” 2021.

Figures and Tables

Figure 1: Frequently used emojis according to Emojipedia February 2026 ratings: used as stimuli

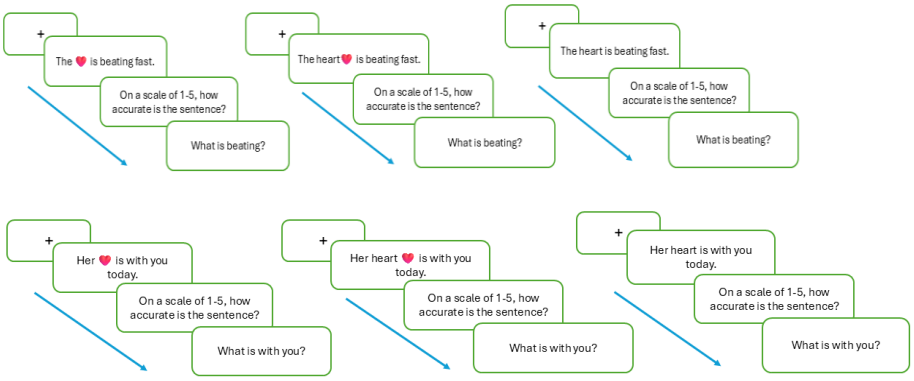

Figure 2: Design and Setup of the experiment; **Top:** Literal Context, **Down:** Metaphorical context.

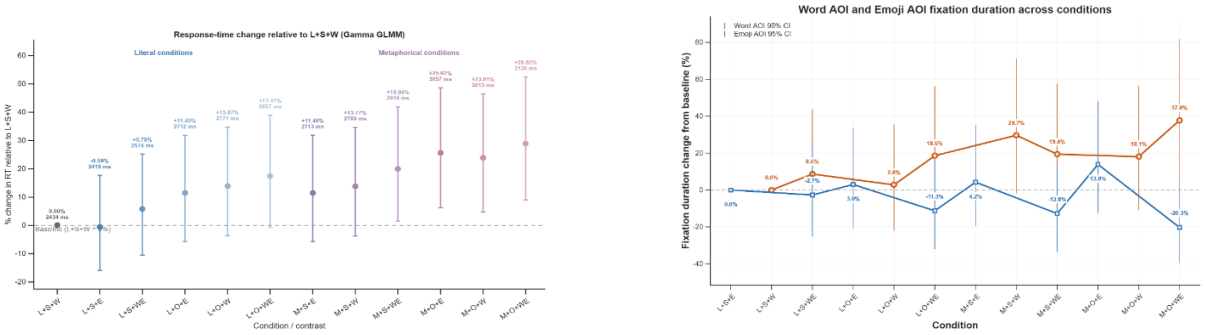

Figure 3 (left): Percentage change across the response times of the different sentence categories. 4 (right): An overview of increase in fixation durations across word and emoji AOI regions

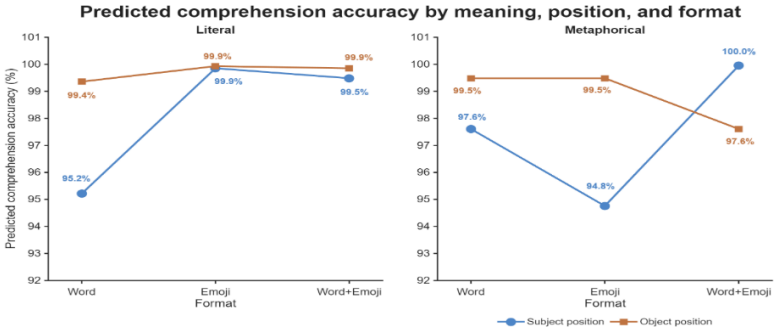

Figure 5: Predicted accuracy levels across conditions after applying Binomial GLMM