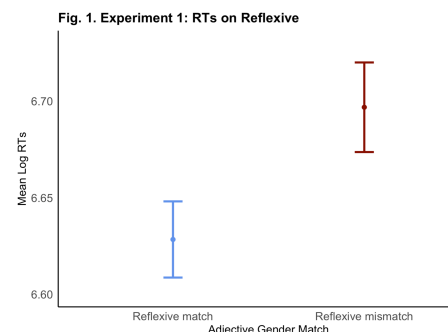


Adjective Gender Stereotypes Shape Coreferential Dependency Formation

Previous research has found that gender bias impacts coreferential dependency formation. When a grammatically accessible noun phrase is judged by the comprehender as “mismatching” a pronoun in gender, such as in *The plumber injured herself*, a slowdown in reading times (RTs) is observed on the reflexive [e.g., 1,2]. This has been termed Gender Mismatch Effect (GMME). A growing body of corpus research [3-5] demonstrates that English adjectives also carry gender stereotypes (e.g., beautiful, gossipy → feminine; handsome, chivalrous → masculine). However, no prior work has investigated if adjective gender biases influence real-time coreference resolution. We address this gap by testing whether gender-biased adjectives modulate processing of reflexives (*herself/himself*). Further, this work investigates where these adjective gender inferences originate. Adjectives have the ability to appear in different informational structural positions and previous work had found that an adjective's interpretation can be sensitive to its information structure [6-8]. For example, prior work shows that adjective subjectivity varies with discourse status: adjectives embedded in non-at-issue or backgrounded positions receive lower subjectivity ratings than those presented as new or at-issue information (Attributive < Appositive Relative Clause (ARC) < Predicative) [7]. In sentence processing, number agreement attraction effects are also found to be sensitive to information structure, occurring in Restrictive Relative Clauses (RRCs) but not in ARCs [9-11]. These findings converge on the idea that information structure can modulate how linguistic features are encoded and subsequently retrieved in comprehension. However, it is important to note that reflexive coreference behaves differently from number agreement. [12] show that reflexives do not exhibit agreement attraction or intrusion effects, arguing for a memory retrieval mechanism that uses primarily syntactic information (e.g., binding principles) to guide retrievals for the reflexive's antecedent. If reflexive binding is truly resistant to agreement attraction and intrusion effects, then adjective gender that modifies the reflexive's antecedent, regardless of that adjective's discourse status, should always remain available in working memory during processing. This raises a critical question for gendered adjectives: Do adjective-based gender expectations pattern like number agreement attraction (susceptible to information structure modulation in processing), or like stable semantic features activated when the adjective is retrieved (and thus impact reflexive coreference regardless of discourse status)? Two competing hypotheses arise. Under a structure-based hypothesis, adjective gender biases arise via contextual inference and should have less impact on reflexive coreference when the adjective is backgrounded or not-at-issue (e.g., in attributive or ARC positions), paralleling the pattern seen in agreement attraction. Under a lexically based hypothesis, adjective gender stereotypes are encoded as stable semantic features, activated whenever the adjective is processed; therefore, they should similarly influence reflexive coreference across all information-structural environments, following the stability of reflexive binding mechanisms. We test these hypotheses in processing by manipulating adjective information structure in several environments (attributive, predicative, RRC, ARC).

Exp 1: Establishing Adjective GMME. Participants (N=50) saw items word-by-word using the G-Maze task[13] in which they needed to distinguish the correct continuation from a distractor. Exp 1 had a 2x2 design: Adjective (masc vs. fem biased) and Gender Match (match vs. mismatch; *herself/himself*), see (1). Gendered adjectives and neutral nouns were selected from a separate norming study we conducted, where participants rated words on a 7-point Likert scale with 1 being “more likely to be a man” and 7 being “more likely to be a woman.” The gendered adjective was always in the attributive position describing a neutral noun.

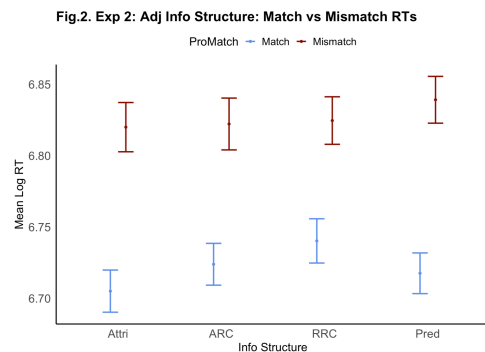


(1) a. Apparently the {*pretty*_{FEM}/*handsome*_{MASC}} employee gained many clients after dedicating {*herself/himself*}

We found a main effect of Gender Match ($p < .001^{***}$) at the reflexive region: RTs were significantly slower when the adjective gender bias mismatched the reflexive – see Fig. 1. These findings show that comprehenders are sensitive to adjective gender bias during coreferential processing.

Exp 2: Information Structure. Exp 2 (150 participants) also used G-maze with a 2 x 4 design: Gender Match (match vs. mismatch with himself/herself reflexives) and Info. Structure (attributive vs. ARC vs. RRC vs. predicative), see (2).

- (2) a. The *pretty* employee gained many clients after dedicating {*herself/himself*}...
 b. The employee, who was *pretty*, gained many clients after dedicating {*herself/himself*}...
 c. The employee who was *pretty* gained many clients after dedicating {*herself/himself*}...
 d. The employee was *pretty* and gained many clients after dedicating {*herself/himself*}...



Adjectives were half feminine-biased and half masculine-biased, and half aesthetic (e.g., *pretty*) and half behavioral (e.g., *chivalrous*). The latter distinction was motivated by prior work showing that aesthetic adjectives are more subjective/multidimensional and thus should be more susceptible to information structure [8, 14]. While we found a significant main effect of Gender Match ($p < .001^{***}$), we did not find a significant Gender Match \times Inf. Structure interaction (all $ps > .10$) – see Fig. 2. A post-hoc exploratory analysis did show that aesthetic adjectives displayed a numerical trend broadly consistent with a structure-based hypothesis.

Exp 3: ARC vs. RRC with aesthetic adjectives. Given the exploratory pattern in Exp 2 and the prevalence of the ARC/RRC contrast in sentence-processing research, Exp 3 (G-maze, 60 participants) focuses on aesthetic adjectives in ARC vs. RRC positions. We still found a significant main effect of Gender Match ($p < .001^{***}$), but no significant interaction of Gender Match \times Info. Structure. Even when focusing exclusively on the adjective type most likely to reveal info structure-based sensitivity, the GMME remained stable across clause types.

Implications. Across three experiments, we provide robust evidence that English adjectives with gender stereotypes influence real-time reflexive resolution. These findings extend earlier corpus-based claims about adjective gender to incremental sentence processing. Crucially, we find no evidence that information structure modulates the GMME. Although subjective adjectives are known to be more context-dependent [7], stereotype-based gender information derived from adjectives is remarkably resilient across structural environments. These findings echo [12], as unlike morphological number agreement features, semantic gender cues remain influential on reflexive coreference even when the adjective appears in a backgrounded, non-at-issue, or parenthetical position. Together, these results support a lexically-based hypothesis: gender biases are part of adjectives' semantic representation and are automatically retrieved and held in working memory when their lexical entry is accessed. Future work will examine cases where the noun itself bears gender bias, allowing adjective-noun competition (e.g., *The nurse who was muscular ... himself/herself*), to test whether semantic competitors can create interference effects under reflexive binding, as this is a scenario structurally unavailable in the present design with neutral nouns.

References

- [1] Carrieras et al., 1996 [2] Sturt, 2003 [3] Moon, 2014 [4] Turrentine et al., 2019 [5] Raphael, 2023 [6] Potts, 2005 [7] Kaiser & Wang 2021 [8] Glass, 2024 [9] Wagers, Lau, & Phillips, 2009 [10] Ng & Husband, 2017 [11] Kim & Xiang 2024 [12] Dillon et al., 2013 [13] Boyce et al., 2020 [14] McNally & Stojanovic, 2017