Prejudice and performativity: can slurs be used without being mentioned?

Background Different types of content behave differently under ellipsis (see, e.g., Esipova 2019 for a brief overview and further references). E.g., presuppositions of predicates encoding a stage of an event (*start*, *stop*...) or factive predicates (*know*, *regret*...) are never ignored under ellipsis while contributions of pure expressives always are:

(1) Pam stopped smoking, {but Kim didn't / and Kim did, too / and so did Kim}.
    → {Pam / Kim} used to smoke.

(2) A: Did you bring a fucking gun to my house?
    B: No, I didn't. / Yes, I did. / Yes, I did so. / Yes, I brought one.
    {→ A / ↛ B} is experiencing strong emotions.

There are a few differences between these two cases: (i) the target piece of meaning in (1) is contributed by the main root of the antecedent VP, but the one in (2) is contributed by an adjunct; (ii) the presupposition of *stop* is a precondition for the antecedent in (1) to make sense, unlike the contribution of *fucking* in (2); and perhaps most importantly, (iii) acts of producing expressives like *fucking* are purely performative, i.e., the speaker achieves their goal (here, expressing their emotions) by virtue of producing a certain form (use via mention), and there is no way to achieve this goal without performing this act (no use without mention).

   The prejudice component of slurs is a more complex case: (i) a slur can be the main root of the antecedent constituent targeted by different types of ellipsis (like *stop*, unlike *fucking*); (ii) however, the prejudice component is not crucial for the sentence containing a slur to make sense (unlike *stop*, like *fucking*); (iii) slurs can be used performatively (use via mention) and can even have a performative effect of offense by virtue of being uttered in the absence of such intent on the speaker's part (mention without use), but it is unclear **if the prejudice component can be preserved if a slur is recovered but not uttered** (use without mention). Saab (2020) suggests that it isn't preserved: "ellipsis is an apt strategy to nullify the bias encoded in some lexical items". However, his conclusion was based on categorically presented introspection-based data only (from Spanish). I investigated this question experimentally in a way that allowed for gradient judgements, by looking at paradigms like (3) (the exchanges are set in a fictional universe where humans co-exist with centaurs, dwarves, elves, orcs, etc. and happen in the context of a criminal investigation):

(3) a. *Context: 'Tusky' is a slur for orcs.*
       Detective: Did you see a tusky?
       Witness: Yes. ('Bare') / Yes, I did. ('VPE') / Yes, I saw one. ('One') / Yes, I saw a tusky. ('Slur') / Yes, I saw an orc. ('Nonslur')
    b. *Context: 'Tusky' is a slur for orcs. This slur can also be used as a verb meaning 'to crawl' (for any race), because orcs are stereotyped as living in caves and, thus, having to crawl through narrow spaces all the time. The detective is asking a question about a human.*
       Detective: What happened next? Did he tusky under the table?
       Witness: Yes. ('Bare') / Yes, he did. ('VPE') / Yes, he did so. ('So') / Yes, he tuskied under the table. ('Slur') / Yes, he crawled under the table. ('Nonslur')
    Question: How likely do you think that this witness is prejudiced against orcs?

Hypothesis I hypothesized that the "prejudice likelihood" inferred from responses like those in (3) is affected by several syntactic, semantic, and pragmatic factors: 1. Maintaining that slurs do have performative effects, I expected the likelihood to be highest when the witness utters the slur

themselves ('Slur'), ostensibly both using and mentioning it. 2. I expected the likelihood to be lowest when the witness tacitly challenges the detective's use of the slur by using the neutral term instead ('Nonslur') in an attempt to minimize complicity (see, e.g., Cepollaro 2020 and references therein on unchallenged slurs). 3. Finally, I hypothesized that the prejudice component of slurs is not exclusively performative. Thus, when the slur is obligatorily recovered, e.g., when it is the main root of the constituent targeted by a proform (more obviously in 'One' for nouns; less obviously in 'VPE' and 'So' for verbs), the prejudice likelihood is expected to be higher than in elliptical responses where the slur is not necessarily recovered ('Bare' and 'VPE' for nouns; 'Bare' for verbs). So, if all parts of the hypothesis are correct, we expect the following picture:

(4) Predicted prejudice likelihood ratings (from lowest to highest)
    a. Nouns: 'Nonslur' < 'Bare'/'VPE' < 'One' < 'Slur'
    b. Verbs: 'Nonslur' < 'Bare' < 'VPE'/'So' < 'Slur'

Methods The experiment involved 10 conditions (2 parts of speech, 5 response types for each). Each participant saw 2 trials per condition and 2 attention checks (22 trials total); the trials looked similarly to (3). Participants assessed the prejudice likelihood by dragging a slider on a pseudo-continuous scale (mapped to 0–100) from 'Not at all likely' to 'Very likely'. Participants were recruited on Prolific (final $N = 128$) and paid £1.25 for completing the task.

Results The results are visualized in Fig. 1. The statistically significant contrasts fully matched the prediction in (4a) for nouns, but only partially matched the prediction in (4b) for verbs:

(5) Statistically significant contrasts in prejudice likelihood ratings (from lowest to highest)
    a. Nouns: 'Nonslur' < 'Bare'/'VPE' < 'One' < 'Slur'
    b. Verbs: 'Nonslur' < 'Bare'/'VPE'/'So' < 'Slur'

Discussion The results for noun slurs corroborate all parts of the hypothesis, suggesting that the prejudice component of slurs is performative, but not exclusively so. This goes contra Saab's categorical claim above (the "bias" is weakened, but not "nullified" when the slur is recovered during ellipsis resolution) and calls for a hybrid analysis for slurs that doesn't reduce their prejudice component to just a presupposition (as in Schlenker 2007) or just a performative effect on the context (as in Potts 2007). The results for verb slurs corroborated parts 1 and 2, but not 3 of the hypothesis, possibly because: (i) in the absence of perfect English counterparts, verb slurs were harder to intuit about, (ii) due to (i) and the less direct link between the meaning of a verb slur and the targeted group, the contrasts were overall less pronounced, and (iii) the identity requirements for verbs in VPE and *do so*-replacement are less clear than for nouns in *one*-replacement.
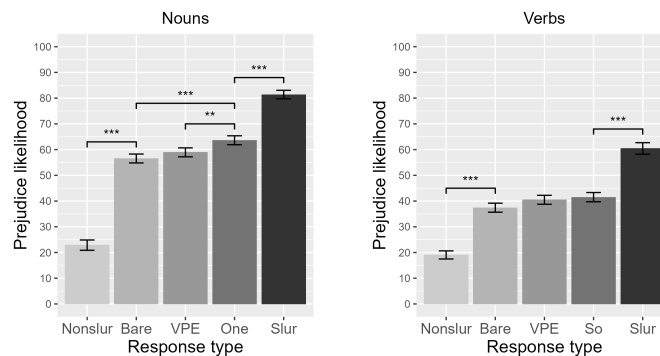

Fig. 1: Bar charts showing mean prejudice likelihood ratings of different types of responses to antecedent utterances with noun and verb slurs, with SE and key significant contrasts indicated.

References

Cepollaro, Bianca. 2020. *Slurs and thick terms: When language encodes values*. Lexington Books.

Esipova, Maria. 2019. *Composition and projection in speech and gesture*: New York University dissertation. URL `https://ling.auf.net/lingbuzz/004676`.

Potts, Christopher. 2007. The expressive dimension. *Theoretical linguistics* 33(2). 165–198. doi: 10.1515/TL.2007.011.

Saab, Andrés. 2020. On the locus of expressivity. Deriving parallel meaning dimensions from architectural considerations. In Eleonora Orlando & Andrés Saab (eds.), *Slurs and expressivity*, 17–44. Lexington Books.

Schlenker, Philippe. 2007. Expressive presuppositions. *Theoretical Linguistics* 33(2). 237–245. doi: 10.1515/TL.2007.017.