Croatian Journal of Philosophy
Vol. XXV, No. 75, 2025
https://doi.org/10.52685/cjp.25.75.4
Received: July 28, 2025
Accepted: September 24, 2025

A Defense of Lexical Accounts of Slurs: Comments on Stojnić and Lepore's Inflammatory Language

LOUISE ANTONY
University of Massachusetts, Amherst, USA and Rutgers University, New Brunswick, USA

I defend a lexical account of slurs against criticisms mounted by Stojnić and Lepore, and present positive reasons for preferring a lexical account over the articulation account they put forward. A lexical account, I argue, explains why articulations of a slur give offense: viz., they are recognized as articulations of a particular word. A lexical account also does better than the articulation account in explaining the naive acquisition and use of slurs, facts about mishearing and mispronouncing, the evolution of certain words into or away from status as slurs, and the fact that "reclaimed" slurring words still give offense when used as slurs.

Keywords: Norms; normativity; functions; social roles; social construction; social kinds; convention.

I admire this book enormously; I cannot claim to have fully digested it. I completely agree with Stojnić's and Lepore's criticisms of content views of slurs, partly just because of my background bias toward externalism in semantics. So I'm prepared to treat slurs as simply meaning their denotations, just as neutral counterpart terms do (or at least, to the same extent as they do), and to look to non-semantic factors to explain the phenomena. And I also agree with the critiques they offer of expressivist and other views that appeal to speaker intentions. (I do have sympathy with the idea (Bolinger, Nunberg, Kirk-Gianni) that a speaker's *choice* (or presumed choice) to use a slurring expression carries information about the speaker's intentions, attitudes or affiliations, particularly if there is a neutral counterpart term available but unused, but I think that word choice in general can carry such information, so there is nothing here that is specific to slurs.)

Another reason I have for rejecting both semantic and expressivist views of slurs – and this reason is something not fully discussed by Stojnić and Lepore – is the fact that slurs can be acquired and used *naively*. That is, the terms can be picked up and used by someone who does not know that the term is a slur. A person can acquire a slurring term without knowing the stereotype associated with the term (if there is one), and without knowing even that the term is, can be used to, or is taken to express contempt. Here's an example: a good while back, when my husband and I lived in Raleigh, N. C., the two of us were chatting with our neighbor, a native Southerner. He had just started a career as a realtor, and we asked him how things were going. "Great," he said. "I just made a terrific deal – the guy wanted $n, but I managed to Jew him down." My husband is Jewish; my neighbor knew that. But it was clear from the immediate change in his expression as he spoke that he had not until that moment connected the expression he was using with Jewish people. He apologized profusely and insisted that that was "just what we called it." My husband and I believed him. After all, he and I had both grown up using the expressions "gyp" (which I would have spelled "J-I-P") and "Paddy wagon" without the least awareness that these were slurs against (respectively) the Romani and the Irish peoples.

Here's a slightly different case, but one which also illustrates naïve acquisition: one summer while I was in graduate school, I got a job as a counselor at a day camp in Dorchester, MA, a city that in the 1970's and 80's had seen a great deal of racial strife. The camp was multi-racial and the administrators were trying valiantly to foster inter-racial friendships. My charges were 8-yr.-olds, but they had already picked up some nasty terms. When some of the Black kids got mad at me, they would, with semantic appropriateness, call me a "honky." When some of the white kids got mad at me, however, they would hurl the N-word at me. What I think was going on was that the white kids (and maybe the Black kids, too) had naively mistaken a racially specific slurring term for a general pejorative.

One more – possible – example: Mark Twain's novel *Huckleberry Finn* is written in the voice of its main character, a 13 (or so) year-old boy in antebellum Missouri. The novel is now notorious because of Twain's use of the N-word in rendering the thoughts and speech of its characters. Twain tells his reader, in an explanatory forenote, that:

> a number of dialects are used, to wit: the Missouri negro dialect; the extremest form of the backwoods Southwestern dialect; the ordinary "Pike County" dialect; and four modified varieties of this last. The shadings have not been done in a haphazard fashion, or by guesswork; but painstakingly, and with the trustworthy guidance and support of personal familiarity with these several forms of speech. (Twain 1884)

(Twain's purpose in adding this note is not, apparently, to apologize for his use of the N-word, but to explain the differences in his various characters' patterns of speech to the "many readers" who, he feared, "would

[without the note] suppose that all these characters were trying to talk alike and not succeeding.") If Twain's research can be trusted, the term he used and put into his characters' voices was *the* term for African-Americans at the time and place, and among the populations, in which the novel is set. If that's the case, then Huck would presumably have acquired the term simply as the common name for Black people, the way liberals and "well-bred" bigots of my parents' generation picked up the terms "Negro" and "Colored" as names for Black people.

Indeed, the move from "Negro" to "Black" as the preferred term for Black people in the U.S. supports, I think, the lexical view. Brian Palmer of *Slate* explains how the terms "Negro" and "colored" moved from being acceptable and even preferred, to being strongly disfavored. Palmer marks the turning point to be the historic speech by Stokely Carmichael in Mississippi in 1966, in which he coined the expression "black power." According to Palmer, Carmichael, in his book *Black Power: The Politics of Liberation in America*, "persuasively argued that the term ["Negro"] implied black inferiority."

Palmer continues:

> Prominent black publications like *Ebony* switched from *Negro* to *black* at the end of the decade, and the masses soon followed. According to a 1968 *Newsweek* poll, more than two-thirds of black Americans still preferred *Negro*, but *black* had become the majority preference by 1974. Both the Associated Press and the *New York Times* abandoned *Negro* in the 1970s, and by the mid-1980s, even the most hidebound institutions, like the U.S. Supreme Court, had largely stopped using *Negro*. (Palmer 2010)

This progression is difficult to understand on any semantic or expressivist view. And neither, I think, can the Articulation Account make sense of it, especially since the N-word was, in the dialects of at least many Americans, *an articulation* of the "respectful" word "Negro."

To be clear, the point I am making here is *linguistic*, not moral. I am not saying that the fact that a slur has been acquired naively makes a person who deploys it is blameless. As is recognized by everyone in this debate, a slur can cause pain whatever the intentions of its speaker, and at a certain point in social evolution, persons become responsible for knowing such matters as that certain expressions are slurs. Rather, the point is that even if the meaning of a slur is something derogatory, one *can* use the slur without knowing that. This is a consequence of externalism, and what Putnam called the "social division of linguistic labor:" we pick up terms from the others in our milieu without needing to know what the others know about them. We certainly do not need, in general, to know the etiology of an expression in order to use it; neither do we need to know how and why a slur counts as a slur. Even had I known that "gyp" derived from "gypsy," that wouldn't have told me that the term was a slur, because I didn't know that the term was a derogatory term for the Roma, nor that these people were derogated as cheaters and thieves.

But to return to Stojnić and Lepore's own discussion. Although I agree with their criticisms of semantic and expressivist views, I am satisfied with an approach to slurs that they reject, namely the *lexical* account. What makes a term a slurring term, I submit, is the fact that the term has a certain kind of history. I'm not sure exactly how to characterize this kind of history precisely, but typically, it has these features: it is, first of all, a name for a group that the members of that group did not give themselves, and secondly, the name was given in the context of oppression, discrimination, or domination. It is also typically the case that the term was and is used by individuals who endorse the oppressive relationship or who advocate (or at least would favor) restoring the oppressive relationship. Slurring terms are offensive to those to whom the terms "apply" because the terms are recognized as having this sort of history. (Uses of the terms by those to whom they apply – ingroup slang or "reclaimed" uses – have to be dealt with separately, as everyone acknowledges.) And the offense occurs because of the history of the term, and often regardless of the user's specific intentions or attitudes. The *potency* of a slur has much to do with the extent to which the oppressive condition still exists, or the extent to which there are people who resent or want to undo social progress.

I was convinced of this much by Luvell Anderson and (an earlier person-stage of) Ernest Lepore (Anderson and Lepore 2011: 25–48). Anderson and Lepore went on, however, to say that slurring words were words that had been *prohibited* – rendered taboo – by some set of authorized authorities. Stojnić and Lepore take this element of the view to be central (as indeed, Anderson and Lepore do) and accordingly criticize "Prohibitionism," and raise several important issues for it. But what about the part of the Anderson-Lepore view that simply says that to be a slur is to be a *word* with a certain history and sociology? One can endorse this part without taking the prohibitionism on board. So, apart from the prohibitionism, what's wrong with a lexical account?

Stojnić and Lepore argue that it is not *words per se* that give offense – it is only *articulations* of words that give offense. They write: "an offense is provoked by specific vehicles for presenting slurs, *not* by occurrences of slurs themselves" (Stojnić and Lepore 2021: 747). I am very puzzled by this distinction. I grant that *no* word has any causal power in itself – words are presumably abstract objects of some sort, and abstract objects have to get themselves realized or materially instantiated in some way or other in order to have any causal effect. Stojnić and Lepore deny that words are abstract objects; they say that they are "shapes." I am very puzzled by this, since I think of shapes as abstract objects, too. But I don't think it's important to settle the metaphysics of words. As long as words are distinct from their individual articulations, I want to go with the words. I contend that the *effects* of the articulations of slurs that S &L make central to their account are themselves

effects of the *recognition* that a given articulation is *an* articulation *of a certain word*. (More on this model in a minute.)

I fully agree with Stojnić and Lepore that it is centrally characteristic of slurs that their articulations evoke negative and often quite painful reactions on the part of auditors of the slur, particularly if the auditor is a member of the group targeted by the slur. But the Articulation Account does not explain *why* articulations of slurs trigger negative reactions – why should the articulation of *these* words have these effects? The explanation certainly seems to be – somehow -- a matter of the history and sociology of the word. Stojnić and Lepore seem to think, however, that somehow the articulations themselves, independent of the words that are expressed, do all the work. "[C]ertain articulations automatically trigger negative associations, which come to be associated with these articulations through a web of complex causal, socio-historical, cultural, and psychological factors." My point, though, is that the Articulation Account offers no explanation for why *articulations* of a word should trigger negative associations, if not because they are articulations of offensive *words*. It's not like the sound or the orthography are inherently offensive, or have, independently of our identifying them as articulations of certain words, negative valence.

Stojnić and Lepore, consider, but do not accept, Mandelbaum and Young's suggestion that there is something inherently negative or distasteful in certain "phonesthetic" features (specifically, "velar plosives") and that the over-representation of such features in slurs may be significant. But I actually found their argument persuasive: if velar plosives turn out to be a significant factor in explaining why certain words (not necessarily slurs) are "disliked" more than others, then it makes sense that such sounds would show up more frequently in insulting or hurtful expressions. I take it, in other words, that Mandelbaum and Young's hypothesis is friendly to the lexical view.

Stojnić and Lepore reject the picture that I'm painting – they say it's a confusion to think that articulations only play a role in triggering offensive effects "insofar as they help us recognize which word is uttered" (Stojnić and Lepore 2021: 752) because such a picture involves the assumption that there's a multi-stage process – first hear the sounds, then figure out what word was articulated, then take offense. But (a) I don't deny that associations can bypass rational or cognitive processing; and (b) I take it for granted, as Stojnić and Lepore seem not to, that very rapid word recognition might *seem* not to involve cognitive processing even when it does; whether this is so is not a question that reflection on the phenomenology can solve.

In making this criticism, Stojnić and Lepore seem to be focused on cognitively *unmediated* association; this focus is misplaced (and, as we'll see in a minute, has to be shifted in order to account for *reclaiming* of slurs.) We know from critiques of behaviorism that the existence

of an association between stimulus and response does *not* mean that there is no cognitive mediation.[1] The explanation for one-trial extinction can be – generally is – that a necessary cognitive link has been broken – that's why the human being who is shown that the shock machine has been unplugged will have no or reduced galvanic skin response when they sit down at the apparatus.

Moreover, I think that Stojnić and Lepore's account seems to deny that the *cognitive state* of simply learning that someone used a certain slur – even if the slur itself is not articulated – can produce the offensive effects. But I think that, as a matter of fact, this happens. Suppose I say to Hermione: "Draco called you a you-know-what" – Hermione might well be– I say probably would be – triggered in a similar way to the way she would have been if I had articulated the slur. (Can we evaluate such counterfactuals? Why not?)

Let's turn then, to facts about the "reclamation" of slurs. While Stojnić and Lepore earlier seem to presuppose a behavioristic view of associations, their discussion here seems to make the pertinent associations highly cognitive – Stojnić and Lepore say that while negative associations cannot be erased, they can be "exploited or subverted in order to signal camaraderie or solidarity." But while one can simply choose to proudly declare oneself a "slut" or a "bitch" or a "queer," if the terms' potency to offend was really fundamentally a matter of ingrained associations, it's hard to see how reclamation could happen quickly, or voluntarily. I am not, obviously, saying that an association *couldn't* be broken by the adoption of some new propositional attitude (since that would conflict with what I've said about (e.g.) one-trial extinction). But it doesn't follow that any given association can always be broken in this way.

Relatedly, Stojnić and Lepore fail to explain why, even after reclamation by members of the in-group, *articulations* by members of *outgroups* still have the power to trigger offensive effects in members of the in-group. Surely the difference is entirely cognitive – the sound of the articulations of the slurring words could be identical between an ingrouper's pronunciation and an out-grouper's; the difference in effect would lie in the *knowledge* of what the out-grouper was doing.

Stojnić and Lepore say that their account does a much better job than its competitors in explaining certain central phenomena involving slurs. One important phenomenon is the fact that slurs *project*, and do so indiscriminately. It doesn't matter if I embed a slurring expression inside a context of indirect quotation – it still retains (or may do) its

[1] See the masterful review of "anomalies" in the behaviorist literature in Brewer (1974). Brewer distinguishes between the existence of an S-R or operant conditioning regularity, and the behaviorist's *explanation* for the regularity. He argues, compellingly, that even apparently very simple S-R regularities are best explained on a cognitivist model, as is shown by the sort of case of one-trial extinction described above.

stinging effect. But I think that the lexical account does just as well in explaining this, once one grants that words always have their effects *via* their articulation. The mere pronouncing of the slurring word can trigger the same painful reactions in an auditor even if the auditor recognizes the slur is being mentioned rather than used by the speaker.

(I want to just float an alternative – or perhaps additional – explanation, compatible with just about any account of slurs, for why slurs fail to project out of indirect discourse contexts. The fact is that we lack, in colloquial and (especially oral) speech, good mechanisms for indicating *direct* as opposed to indirect quotation. In written speech, we can use quotation marks, and in oral speech, we can say explicitly things like "Draco said, and I quote, ….." But these explicit indicators that words are being mentioned and not used are not obligatory. (So-called "air quotes," in oral speech, are ambiguous between indicating that the words air-quoted are merely being mentioned, and indicating that the speaker is using them in a "so-called" way.) In ordinary contexts, saying or writing "Draco said that Hermione is a mudblood" is simply ambiguous between a reading in which I'm *paraphrasing* Draco and one in which I'm *quoting* Draco. (And then of course there's the possibility that I'm essentially doing both – that is, Draco used the word "mudblood" and I also use that derogatory word to refer to muggles.) I wonder if there would be less evidence of projection if we were always perfectly explicit when we are mentioning the offensive word rather than using it. And even lesser if we, say, offer an explanation and apology in advance: suppose that Draco has beaten up Hermione, and the issue is whether the degree of offense should be raised because it was a hate crime (as can happen in the U.S.). The prosecutor might say: "I've very sorry to have to use the language I'm about to use, because I know that it is offensive to many people, and it is not language that I use myself, but I need you to know exactly what Draco said before he attacked Hermione: 'You are a rotten mudblood.'" I am not suggesting that the occurrence of the slur "mudblood" would not trigger difficult feelings for the persons targeted by the slur, but such feelings would, I submit, be different in nature from the ones triggered by a hateful *use* of the term.)

In further defense of a lexical account, I want to point out some important questions that I think the Articulation Account fails to answer, illustrated by the following anecdote, recounted in the British paper *The Independent*. The matter involved the philosopher Sidney Morgenbesser – renowned for his quick wit as well as his philosophical depth:

> [An] unfortunate encounter with the police occurred when he lit up his pipe on the way out of a [New York City] subway station. Morgenbesser protested to the officer who tried to stop him that the rules covered smoking in the station, not outside. The cop conceded he had a point, but said: "If I let you get away with it, I'd have to let everyone get away with it." To which Morgenbesser, in a famously misunderstood line, retorted: "Who do you think

> you are, Kant?" Hauled off to the precinct lock-up, Morgenbesser only won his freedom after a colleague showed up and explained the Categorical Imperative to the nonplussed boys in blue. (Gumbel 2004)

One important question this anecdote raises is: what counts as an *articulation of* a slur? Did Morgenbesser call the cop a c*nt, or did he not? Another question: when one is *misunderstood* as having uttered a slur, is that fact enough for the person *to have in fact* uttered a slur? (I'm not asking about the culpability of such a person – I'm only wondering if they've articulated a slur or not.) Note that S & L contend that it is a count against lexical accounts that some non-slurring words have articulations similar enough to common articulations of a slurring term to trigger the same articulations. But I don't see that the lexical account has any problem with this. The fact is that almost any word can be *misheard*, and the fact that it's misheard *as being* an articulation of word X doesn't tell us anything, other than that the articulations of some words are identical or similar to each other.

So I agree that the explanation for why pejorative effects project so indiscriminately has to do with associative effects – although I'll say a little more about the limits of this in a minute – but it's not *just* a matter of association; the effects of the association can be erased in at least some cases where one finds out more. If cop in the Morgenbesser story was capable of realizing that Morgenbesser was invoking the name of the philosopher Immanuel Kant, and wasn't calling him the C-word, he might have just chuckled and said, "Oh, I see – OK." (He might still have been insulted or made angry, but on a different basis.) That is, it might have mattered to the cop *what word* Morgenbesser was articulating. But it would be the same if Morgenbesser had been *misspeaking* – if he actually was trying to call the cop a c*nt, but mispronounced it as "kant" – if the cop understood Morgenbesser as invoking a philosopher, he would probably become insulted if he came to realize what Morgenbesser had been *trying* to say.

Finally, the Articulation account also has – or appears to have – an implausible consequence, namely, that certain speakers might be *unable* to slur. It's a general fact about words that it is through their articulations that we come to know *which* words have been articulated. Stojnić and Lepore cite as a virtue of the Articulation Account that "absent certain articulations, slurs do not offend." Well, in general, if a slur is misarticulated, the auditor may well fail to appreciate that a slur has been uttered. And no one can take offense without at least believing that an offense has been committed – or at least attempted. But someone with an uncommon speech pattern (maybe our hypothetical Morgenbesser trying to insult the cop) or someone who is not a native speaker of the language in which a certain word is a slur, may be unable to articulate the slur in a way that a target of the slur can understand. Do Stojnić and Lepore want to say that a person *cannot commit* the slur that they are trying to commit? Consider the following quote:

> Absent a particular articulation that carries negative associations, the tokening loses its sting; and the presence of articulation—the specific acoustic or orthographic shape that matches (or closely resembles) the canonical articulation of a slur on its own can trigger the pejorative potential, even when no word—and so no slur—has been tokened. (Stojnić and Lepore 2025: 108)

But I'm looking at the opposite case: Suppose that someone, say Sam, who Ray intended to insult with the slur, figures out – say, from hearing Ray express certain opinions and intentions (not involving the slurring word), or maybe by being told that that was Ray's intention, that Ray was trying to articulate the slur – I expect that this knowledge would be sufficient for Sam to take offense in exactly the same way and to the same degree as they would have had Ray successfully articulated the slur.

Stojnić and Lepore address this worry – they say: "But here, the speaker *is* still tokening a term: one can badly misarticulate a term while still articulating that term" (Stojnić and Lepore 2025: 110) but I don't see what, on their account warrants this conclusion. I say, it's the fact that they are trying to articulate a slurring *word* that makes them guilty of committing a slur. The usual negative associations are not triggered, yet there was still a slur committed.

They say:

> An expression that starts off as neutral can, over time, acquire negative associations, becoming a slur; unless we think of all such changes as ones in *meaning*, we cannot conclude that articulations trigger pejorative effects only derivatively, by articulating terms with offensive meanings. (Stojnić and Lepore 2025: 112)

I quite agree that there's no reason to view such processes in terms of changes in *meaning*, but that doesn't count against the view that it's the word – because of its history and sociology – that can, over time, become a slur.

In conclusion, let me reiterate that I have enormous respect for this book, and that I learned a great deal about slurs from reading it. Because I have here focused on my disagreements with Stojnić and Lepore, I have not been able to do justice to the richly detailed, original and systematic arguments they make against views with which the three of us disagree, systematizing and illuminating a number of difficult issues about the interaction among semantics, pragmatics, and social psychology. Anyone interested in slurs – or in language in general – needs to read this book.

References

Anderson, L. and Lepore, E. 2011. "Slurring Words." *Noûs* 47 (1): 25–48.

Brewer, W. F. 1974. "There is No Convincing Evidence for Operant or Classical Conditioning in Adult Humans." In W. B. Weimer and D. S. Palermo (eds.). *Cognition and the Symbolic Processes*. Hillsdale: Erlbaum, 1–42.

Gumbel, A. 2004. "Professor Sidney Morgenbesser." *The Independent*. August 5, 2004.

Cited at: https://en.wikipedia.org/wiki/Sidney_Morgenbesser [accessed August 23, 2024]

Palmer, B. 2010. "When Did the Word Negro Become Taboo?" *Slate* January 11, 2010. https://slate.com/news-and-politics/2010/01/how-old-was-harry-reid-when-the-word-negro-became-taboo.html Accessed July 16, 2025.

Stojnić, U. and Lepore, E. 2021. "Inescapable articulations: Vessels of lexical effects." *Noûs* 56 (3): 742–760.

Stojnić, U. and Lepore, E. 2025. *Inflammatory Language: Its Linguistics and Philosophy*. Oxford: Oxford University Press.

Twain, M. 1884. "Explanatory." *The Adventures of Huckleberry Finn*. Glassbook Classic, https://contentserver.adobe.com/store/books/HuckFinn.pdf