

(How) Does Pronoun Gender Matter?

Audrey Devine-Eller

Ahearn, Language as Social Action, Fall 2006

12/15/2006

In this paper, I examine data I have collected over the past several years about the use of gendered pronouns in English. I am interested in what happens when people use and hear gendered pronouns and terms – that is, what difference does it make if someone uses “him” versus “her”? What about neutral terms, such as avoiding pronouns altogether or using “they” as a singular generic? Specifically, I want to know: Does the gender of the pronoun used in SAT Verbal questions affect performance on test questions? And if so, whose performance, why or how?

Research on English gender pronouns is popular right now from within two different, often separate, disciplinary tracks. The first line of research about gender pronouns occurred with the second wave of feminism, mostly from within sociology, in the seventies and early 80s. Feminists claimed that using generic-as-masculine pronouns (that is, using “he” to refer to any person, male or female), conceptually excludes women from public domains. The famous 1973 text-illustration experiment showed that students did indeed disproportionately choose male illustrations for sections titled “Social Man” than for sections titled “Society”- for which they chose more gender-balanced representations (Schneider and Hacker 1973). Silveira (1980) developed the “people=male” hypothesis, that seeming generics are often interpreted as

masculine, and further concluded “that a reduction in the generic use of *man* and *he* would result in a long term reduction in sexist thinking” (165).

Sociologists also have conducted research that indicates that gendered pronouns really do ‘make a difference’ in concrete terms. Walsh, Hickey, and Duffy (1999) found that, on standardized math test questions, all students answered more accurately on male-labeled questions than on female-labeled or neutral questions. Guyatt et al. (1997) found that language use can indicate underlying attitudes about gender among medical interns. Parks and Robertson (2004) found that attitudes toward sexist language are mediated by attitudes toward women. That is, we might be able to find out about underlying differences in attitudes by studying the language people use, and vice versa.

My current research falls into this line of inquiry, using questions from the Verbal portion of the SAT. The SAT is a particularly appropriate place to ask these questions, as it has been the focus of much recent controversy and sociological research. The practical importance is clear. In 2005, on the old 200-800 point scale, college-bound boys scored 34 points higher than girls in math, and – despite the stereotype that girls are better at verbal tasks than boys – boys scored 8 points higher in verbal (standard deviations are approximately 115) (College Board 2005). These score-gaps make a difference in terms of the colleges students get admitted to, the financial aid and scholarships that are available, et cetera.

Due¹, I think, to this stereotype about girls doing better on verbal tests, and to fears about girls’ underperformance and underrepresentation in math and science, studies have focused on the gender of characters in math and science textbooks and on

¹ A brief Google search indicates that this language ideology isn’t widely held ☺.

the SAT. They have generally found that students perform better on math and science questions when they identify with the characters in the problem or with the subject matter; the argument is that this is facilitated for girls when problems use female characters.

Nonetheless, there have not been many studies to date that specifically investigate these questions with verbal questions from the SAT, which I do here. If girls do more poorly on math and science questions with masculine pronouns, what happens on verbal questions? And what happens to boys who read questions with feminine pronouns? Finally, I add an emphasis on neutral pronouns and referents. Only a few studies have attempted to theorize *how* people interpret supposedly-neutral referents (such as “the student”). Suzanne Romaine claims that “we do not seem to have strong mental images of androgyny to match gender-neutral terms” (Romaine 1999). This implies that we might interpret seemingly-androgynous terms in accordance with the existing, gendered schema we have.

However, I also wish to situate my study with the Whorfian tradition, which investigates the way our language – especially our grammatical categories – impacts thought and social structure. Even though some of this research cited above – Silveira’s, in particular – was and is explicitly interested in the connections between language use and thought, there has been little theoretical crossover with linguists and linguistic anthropologists who study such links.

The so-called Sapir-Whorf Hypothesis posits some kind of connection between our language, our thought, and our reality. Sapir writes in a famous passage:

"It is quite an illusion to imagine that one adjusts to reality without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that the 'real world' is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached" (*The Status of Linguistics as a Science*, 162).

This rather radically focuses our attention away from the representational functions of language, and reminds us that reality is always constructed in and through social interaction, including through language.

I follow Hill and Mannheim (1992) in treating the so-called Sapir-Whorf Hypothesis not as a hypothesis but as an axiom, "part of the initial epistemology and methodology of the linguistic anthropologist" (383). Neither Whorf nor Sapir proposed any kind of falsifiable hypothesis, but instead took as a founding principle that there is a (fuzzy) relationship among language, thought, and meaning. This is the claim: "that grammatical categories, to the extent that they are obligatory or habitual, and relatively inaccessible to the average speaker's consciousness, will form a privileged location for transmitting and reproducing cultural and social categories" (Hill and Mannheim 387). English gender pronouns, to me and to Hill and Mannheim and many others, are one clear place to search for such "Whorfian" effects. As cited above in the sociological research, we can see how "the structure of a system of grammatical categories affects the social ontology posited by the speakers" (389).

John Lucy, a linguist and psychologist, takes seriously the call for experimental evidence about Whorfian effects. He is critical of much contemporary psychological research, which he says largely misses the point. In a (1996) review of the empirical evidence for linguistic relativity, Lucy calls for more research that works cross-linguistically (so as not to privilege one language's categories), deals with a central

grammatical aspect of language (not just a small set of lexical items, such as color terms), provides direct evidence regarding individual cognition, and shows that patterns take place in everyday behavior. While my current research does not meet all of these criteria (it does not work cross-linguistically, and does not study everyday behavior), I hope to show that it is still a worthwhile contribution to this literature. In particular, Lucy calls for researchers to “examine to what extent culturally specific patterns of use – both beliefs and practices – mediate the impact of language structure on thought or have their own direct effects independent of structural type” (37). I argue that my research fits here: in addition to studying performance outcomes on standardized test questions, I conducted focus groups to ask people how they thought and felt about gender pronouns on those standardized tests (see below for details).

Thus, I see the primary contribution of my work to the literature on language ideologies, especially as language ideologies might mediate the effects of linguistic relativity. The study of language ideologies began in earnest with a 1994 conference (Kroskrity 2000) that attempted to reclaim as data not just what people *do* with language, but what they say about what they do with language. Definitions of language ideologies are contested, but I am partial to Irvine’s (cited in Kroskrity 2000): “the cultural system of ideas about social and linguistic relationships, together with their loading of moral and political interests.” Thus, language ideologies are a mediating layer between social structures and forms of talk. Kroskrity reminds us that language ideologies should be conceived of as multiple, and as imbued with power relations that index or construct social hegemonies.

The survey research

I prepared a survey that included 10 sentence completion questions from the SAT. I administered these surveys in Introduction to Sociology (101) classes at Rutgers University in the 2004-2005 academic year, and obtained a total of 1196 completed surveys. (This is approximately a 60% response rate of enrolled students in classes I visited, or 53% of all enrolled students; see appendix for descriptive statistics on self-reported demographic measurements.) Most students were about 20 years old, and nearly half were non-white.

The important outcome variable is the student's score, out of 10, on the sentence completions. I randomly distributed different versions of these sentence completions to students. The questions varied by the pronoun used – for example, one version included all feminine pronouns and referents, one all masculine, and one all neutral.

Sample Question²:

Feminine: Unable to decide between a career in biology and one in philosophy, **Gwen** ----- **her** two interests and became a medical ethicist.

Masculine: Unable to decide between a career in biology and one in philosophy, **Gary** ----- **his** two interests and became a medical ethicist.

Neutral: Unable to decide between a career in biology and one in philosophy, **the student** ----- **the** two interests and became a medical ethicist.

Answers: (A) reclaimed **(B) merged** (C) defined (D) abandoned (E) conveyed

In addition, I conducted 3 focus groups with students who had previously taken this survey in a class. In the focus group, students discussed whether they thought the different pronouns would make a difference, why or why not, and in what directions.

² SAT Test questions selected from 10 Real SATs, College Board, 2003. Reprinted by permission of the College Board, the copyright owner.

Below, I will briefly recount my statistical findings from the survey, but will focus the rest of the paper on the focus group data.

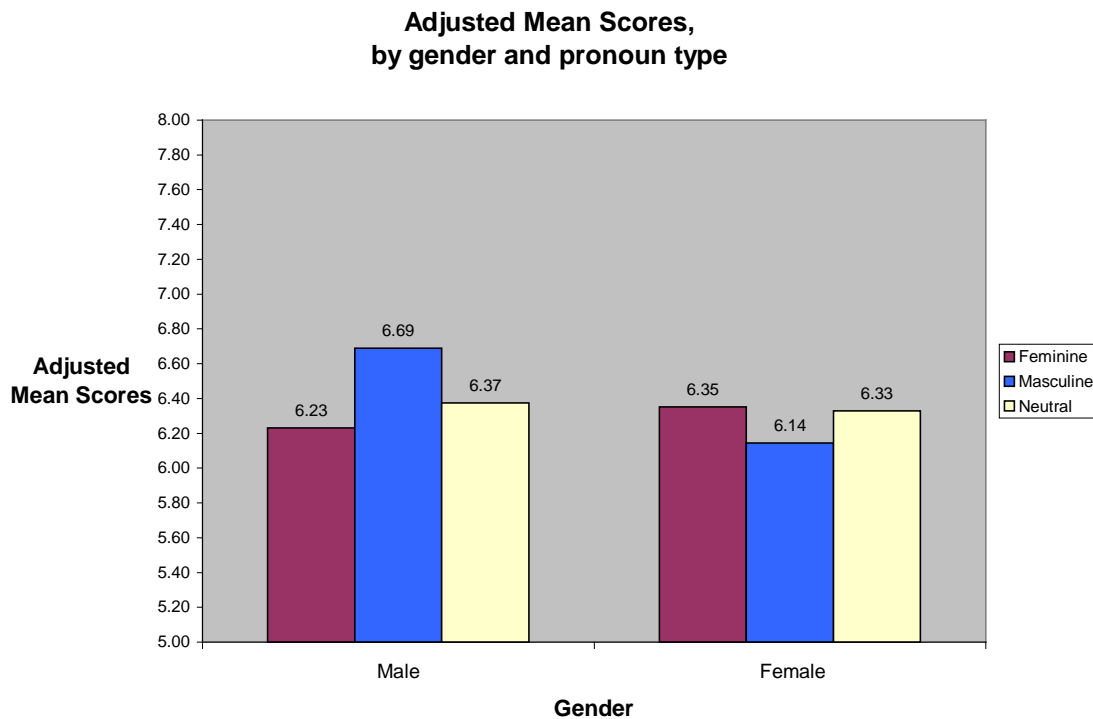
Using two-group *t*-tests, I compared six groups (3 pronouns by 2 genders of test-taker). I found two things of note:

1. that men performed significantly better on the masculine test than women did ($p=.019$)
2. that men performed significantly better on the masculine test than they did on the feminine test ($p=.032$)

I next performed linear regressions, using the demographic characteristics of students – sex, age, year in school, race, SAT scores, SAT prep, first language, and parents' education (See Figure 1 for adjusted mean scores; see appendix for full models). First, as one might expect, SAT scores are highly predictive of the score on my verbal test. For every additional point on the SAT, students scored .396 standard deviations higher on my test. Race is significant too: non-whites scored significantly worse than whites, as did students whose first language is NOT English. These results won't surprise anyone familiar with scoring trends on the SAT. In addition, the interaction term of masculine pronoun*female test-taker is significant. This indicates that there is a significant negative effect for women taking the masculine-pronoun test – an effect equal in size to the negative effect of having English as a second language, and larger than the race effect!

The most advantageous combination is men taking the masculine pronoun test. The most disadvantageous condition is men taking the feminine pronoun test: this actually lowers their scores. The scores for the women all fall in the middle, with the neutral and feminine conditions resulting in higher scores than the masculine. Notice, especially, that women did better on the feminine and worse on the masculine test – exactly the opposite of how men did.

Figure 1.



This means that there are differential effects by gender. This general result is unexpected for several reasons. Previous research focused on the finding that women perform better on feminine-pronoun/character test questions. My study supports this, but the striking thing here is not how well women performed on feminine test questions,

but how poorly they do on masculine test questions and how well men do on those masculine test questions. All other scores are in the middle range between these extremes.

Turning to the linguistic relativity literature, this finding is truly strange. This data seems to complicate a traditional Sapir-Whorfian hypothesis about the influence of grammatical categories on thought. If this is a relativity effect, how are we to explain differential effects for native speakers of the “same” language? Sapir wrote that “The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached”. This implies that, within one language, speakers will share one homogeneous worldview. This is problematic for understanding how men and women could process gender pronouns differently.

One possible explanation is that men and women do, in fact, speak different languages or come from different societies. This idea was originated by Maltz and Borker (1982), who sought to explain cross-gender miscommunication in terms of cross-cultural miscommunication. Though the idea still sells popular books (such as those of Deborah Tannen), it has been more or less thoroughly discredited by linguistic anthropologists. It seems regressive, then, to sidestep the relativity problem this way.

Instead, I propose that there is some other mechanism that mediates between the grammatical categories and people’s performance on the test. Perhaps, as the earlier sociological studies conclude, it is a simple matter of the test-taker identifying with the question. This reasoning relies on the link that when people identify with a question, they will be more interested, and being more interested causes them to work harder and results in higher accuracy. This seems plausible for men’s scores, as they

did the best on masculine, middling on neutral, and the worst on feminine pronouns. In addition, it might explain women's performance very well: women did equally well on feminine and neutral pronouns, and worse on the masculine questions. However, this requires theorizing that women read neutral pronouns/referents as more inclusive than men do.

Perhaps this phenomenon can best be explained by reference to the schemas individuals use to process information on a test. A schema is

“a knowledge representation that provides an organized set of expectations about a given situation, including the relations of the parts to the whole and to each other. Events, scenes, and stories are types of schematically organized knowledge. Schematic structures are held to be automatically activated in familiar situations and guide comprehension, action, and later recall” (Slackman and Nelson 1984:329, cited in Duranti 1994:60).

Perhaps men are more apt at processing masculine pronouns – both because they more readily identify with them, *and* because masculine pronouns exist on that ‘slippery slope’ between specific and generic, *and* because the masculine generic is more typically associated with formal situations and schooling. Perhaps men are less able to process feminine pronouns because they do not fit a familiar schema.³

In addition, there may be some effect of positionality on the outcome of the test. Specifically, perhaps men and women are differentially aware of the political or practical impact of different pronouns/referents. Recall the Guyatt et al. (1997) study that found that language use can indicate underlying attitudes about gender, and the finding of Parks and Robertson (2004) that attitudes toward sexist language are mediated by

³ Ortner (1989) discusses cultural schemas, which she sees as “preorganized schemes of action” that incorporate multiple ideologies and symbols and are, in Bourdieu's terms, structuring structures. This is a useful way of understanding the group of language ideologies I find among focus group participants, but not a useful (because not “micro”) explanation of why people perform differently on standardized tests. The more psychological theories of schema are more relevant at this level.

attitudes toward women. It is possible that if women are more aware (with Du Boisian double-consciousness?) of gender relations or gendered language, their performance will be affected.⁴

Or, perhaps, as I would like to suggest in this paper, all of the above explanations might be consistent with language ideologies as mediating between gender pronouns and peoples' performances on the test. As Lucy suggested, there may be beliefs and practices that "mediate the impact of language structure on thought or have their own direct effects independent of structural type" (37). It seems plausible to me that language ideologies do precisely this. In fact, I see no reason language ideologies can't explain all of the possibilities I have outlined above. If men identify more with masculine pronouns, and women with feminine, this is perhaps due to a language ideology that connects grammatical gender with human gender. If we process some things faster because they fit into our existing schemas for interpretation, how is this different from saying we have particular expectations stemming from our language ideologies; or at least that our language ideologies and schemas are mutually constitutive? If there is an effect of positionality, then too this might be caused by differing language ideologies. For example, a person who has explicit education about gender pronouns will have different language ideologies and meanings associated with those gender pronouns. I will suggest below that beliefs – in the form of language ideologies – are mediating influences on standardized test performance.

It should be clear that I am not sure these explanations are complete or mutually exclusive, nor that they need to be. They do not necessarily rule each other out. For

⁴ In fact, my focus group participants do argue that certain people are more likely to notice the gender pronouns. They do not tie this to test-taker gender, however, but to the possession of an activist or "women studies" consciousness.

example, it is possible that the schema explanation is the most appropriate from a psychological perspective, but that schemas themselves arise due to language beliefs and practices that rigidify, making some language formats easier and quicker to appropriate. I hope to narrow or refine these possibilities by listening to what people actually say they think is going on. Of course, as Frank Boas would remind us, asking people only gives us access to second order explanations. What people think is going on might have very little to do with what is actually going on. But as the language ideology theorists, like Kroskrity, remind us, *what people think is going on* is itself important ethnographic data. I hope that by *asking* people what they think, I can develop a better theory of what language ideologies are relevant here, and whether or not those ideologies are having an impact on language use and accuracy on verbal skills tests.

The Focus Groups

I conducted three focus groups with between two and five participants in April, 2005. Participants were volunteers who had agreed to be contacted after participating in the survey in their Introduction to Sociology classes. They were induced by offers of pizza and soda. Focus groups lasted between 45 and 75 minutes. Participants had printed copies of the survey in front of them during the focus groups, and they refer to it often. (See Appendix for the SAT questions portion of the survey.) There were, of course, differences among the three focus groups, but their similarities were striking, in terms of the progression of the discussion and the theories the participants developed about gendered pronouns on the survey and in language more generally. I will focus on

one group in the discussion below; this was the largest group and the one that talked the longest.

There were six participants, listed below with self-identified demographics. I asked, both on the consent form and for name tags during the focus group, that participants write down the first name they wished to be called for the purposes of the research; thus names might or might not be their “real” names.

Audrey: Interviewer, Female, 26 years old, white, graduate student, sociology

Lily: Female, 21 years old, Chinese/Hispanic, 2nd year, biochemistry

Idi: Female, 19 years old, Nigerian, 1st year, sociology major

Sami: Female, 22 years old, [Asian] Indian, 4th year, criminal justice

Adi (Aditya): Male, 20 years old, no race listed, 4th year, nutrition

Aaron: Male, 27 years old, white, 3rd year, psychology

The discussions tended to progress in similar ways. I started out by asking participants to try to identify any differences among their SAT questions, and they generally agreed that they had identical versions. For example, Adi, after comparing with neighbors, said, “Everything’s the same it’s just different order as far as we can tell⁵ (116)”. Here, he must be referring to the overall ordering of the survey (including SAT questions, personality questions, and demographics). Only when I had participants read their SAT questions out loud were they able to spot the differences – and even then, only after

⁵ Quotes in the text of this paper are written exactly as participants said them, but I have excluded the minimal positive responses of other participants (usually me, as focus group leader). I only quote in this way when one person has an extended turn that is only overlapped by minimal positive responses. Line numbers are taken from the full transcript (not from the transcript excerpts included in the appendix).

some difficulty. The relevant discussion can be found in Transcript Excerpt 1 (lines 128-175 of full transcript).

In this excerpt, I ask Sami to read her question 1. Aaron immediately says that he does not have the same question. I ask him to read his, and he starts out, “*She took*” – and stops short when he thinks that he does in fact have the same question Sami had – even though he didn’t! In other words, the participants here had a difficult time figuring out what was similar and different about the two sentences. An immediate first reaction that they were different turned into a longer reflective decision that they were the same, accompanied by explanations of why they might have been wrong the first time (e.g., Adi explains that Sami read too quickly). Only after a third look, with Sami re-reading her sentence, did Adi and Aaron realize the gender pronoun differences.

I next asked participants whether they thought the gender pronouns made a difference on the test. There was a great deal of disagreement on this, mostly centering on whether test-takers would be aware of the pronouns while they were answering questions. Sami immediately says she thinks that the gender pronouns were important, and the others initially agreed with her, seeing in the questions “male or female biases” (Aaron) or “gender stereotyping” (Adi). Aaron notes being surprised at the inclusion of Anna Freud.

But the consensus falls apart soon after this point, when Adi proposes that perhaps there is some subconscious bias in effect, but that could be confounded by the “possibility if they’re just dumb” they “wouldn’t be able to really properly answer the question”. Adi talks through how a (presumably “dumb”) person would approach the questions: subconscious perceptions could influence a person to answer in a particular

way, in part apparently influenced by gender pronouns. Sami defends her position, and for the first time brings the discussion to a level beyond the SAT, with the representation of women on juries.

Aaron and Lily, at this point, begin to argue strongly that the gender pronouns do not make a difference on this test, but they might in other contexts. Lily raises the possibility that taking a standardized test is a mechanical process and people are “programmed” to look for key words and to ignore the rest of the question. Adi agrees that this is a possibility for those who have prepared for the SAT, but claims that not everyone has done that kind of studying, so they might approach the SAT in an unprogrammed manner. Lily argues this:

yeah but you could assume that the person that wrote the the S- the test is actually like he might make it different if he put he or she but when the person is writing it he's also doing it like a /program/ too he's also like oh let me open the dictionary and this word and /look at the other/ the meaning and then he just plop he she it a student dog cat like whatever he just like he also programs it like so like when he goes to the /public/ it will work by the same standards like it's not that if I make something like search for your opinion I might look for like I might get a different answer where if I put he or she but if I do it like by I don't know like just the definition or just like you open the dictionary independent of the you put he or she before the verb it's gonna come out the same (751-803)

In other words, Lily argues, whether the student taking the test is acting in a programmed way or not, the questions themselves are written in a programmed way that makes them less likely to be influenced by the gender pronouns used. Idi speaks here for the first time, using Lily's words and supporting her ideas: she says she didn't notice the pronouns, but even if she had, “I wouldn't have placed like any type of like significance cause like I was just like mechanical you know”.

Again, Sami argues for her position that the pronouns do matter, this time making more reference to the test itself and less to the broader context. She says,

right now if you take a look at the LSAT I just took the LSAT like a couple of weeks ago and just to be politically correct they started to add names like Gupta Vargas Smith Williamson Johnson like because the thing is that most people might not think about it you might think okay they're programmed but I mean things like the civil rights movement affirmative action those happened for a reason the people that might be white and might be male that might be female that might be at the top of the hierarchy whether it be race whether it be social class economics I mean the other people lower they feel that crunch they feel that exclusion so now with affirmative action equality they're trying to say ok you guys are in it too you know you don't have to just be white to have your name in the LSAT you don't have to just be black you know you can be Gupta you can be a Vargas you can be female you can be male you know you can be whatever (964-1024)

Here Sami argues that what test-takers see on the test does matter, both for the test itself (as she was arguing earlier) and for the degree to which it reflects the identities of test-takers and the equality that is desired in the broader social context.

This, then, is the major disagreement that arose during all of the focus groups. Some students claim that, in the context of the test, test-taking is a programmed skill and so test-takers are unaware of the gender pronouns used – or, even if they are aware of them, the gender pronouns simply do not matter due to the artificial nature of the test questions. On the other side, some students argue that the identities referenced on the test do in fact impact test-takers, at least in part because they resonate with issues of equality in the broader social context, and that test-takers will perform better on tests in which they feel represented.

A secondary disagreement is the degree to which test-takers are conscious of the gender pronouns in the questions. Most students make draw a causal connection from not noticing (or not taking into account) the gender pronouns to having those gender pronouns not matter. They do this in support of different conclusions, however. Aaron argues that while educated people answer the questions correctly and do not

take the gender pronouns into account, less educated people will have more difficulties discounting the pronouns: “we would just look right past it but perhaps people who are not as educated uh would have would have problems with it they might I mean they’re gonna make errors anyway but they might make even more errors or or biased errors depending on their background” (1120-1136). He thus assumes that for educated people, gender pronoun bias does not matter. He later references the Anna Freud sentence again: an uneducated person, he claims, might be more likely to be surprised or caught off guard or to rely on stereotypes when answering that question⁶. Sami, unsurprisingly, disagrees; she believes that people who are educated are more likely to notice when their identity is not represented, because they have read more about activism and civil rights. Again, however, she implicitly ties this to the claim that when people do not notice, it does not make a difference in their performance.

These are not a merely academic issue for participants in the focus group; they vigorously defended their positions, and while usually polite, challenged each other’s views. Transcript Excerpt 2 (lines 1510-1643 of full transcript) contains a discussion between Sami, Adi, and Lily which takes place well into the focus group. The lines are already clearly drawn. Lily strongly believes that pronouns make no difference on the test, but do make a difference in wider contexts. Sami believes they always make a difference. Throughout the focus group, and also here, Adi acts as a mediator, often agreeing with both sides. Just before this excerpt, Adi challenges Sami’s belief that it matters what names are used on standardized tests. Sami has already been challenged on this point before, so here Adi takes a different tactic. Instead of arguing that she is

⁶ Compare this to a participant in a different focus group, who noted that “the names are kind of white names, Anna Freud or Sigmund Freud”. Aaron noticed Anna Freud because of her gender; this other participant noticed the ethnic identity of the names (like Sami).

wrong, he begins to argue that she is more sensitive to this issue than the average person.

In responding, Sami claims that of course these things matter, and again brings the issue to a broader social context of rental markets and job applications. Adi attempts to interrupt and redirect her to the test but she continues. Note that he appears to be convinced by at least one of her arguments, about changing one's name⁷. This comment might have hit home for Adi: he wrote "Aditya" on the consent/information sheet, but "Adi" on his name tag.

At this point Lily jumps back in and challenges Sami to recall the question(s) that bothered her on the LSAT. Sami says she doesn't remember any specific questions, only the names, and is dismissed by both Adi and Lily. Adi says, "she already speaks like a lawyer," discounting Sami's more complex explanation (she remembers the names, not the questions) as double-speak. Lily relies on a criticism that Adi and others had already brought up: that Sami must be more aware of the names because she was already, in a sense, primed to notice them.

One important thing I would like to reiterate here is the belief, among all the students, that *noticing* something is a prerequisite for *having it matter* – at least on a standardized test⁸. However, remember that the classical statement of the linguistic relativity hypothesis relies on *unconscious*, habitual, obligatory facets of language. From this perspective, Lily's argument that people are doing SAT questions in a programmed,

⁷ Also note that, logically, she must have intended to say "from Samiha to Sami", rather than the other way around, to be parallel to her other statements and as a self-reflexive commentary.

⁸ There is one exception to this; a participant in another focus group thought that pronouns might make more of a difference if the test-taker is not paying attention.

unthinking way actually leaves *more* room for the effects of language on thought. And, in fact, notice that when Lily talks about the programmed way in which people write test questions, she uses only “he/him” to refer to the hypothetical writer. She can be seen in this instance as unconsciously using an obligatory grammatical category which, arguably, would impact others’ perceptions of the hypothetical test writer.

The linguistic relativity axiom does not, however, necessarily prevent people from becoming self-reflective about their language use and thought categories. Sami’s claim rings true, that people who are more educated will be more aware of inequities in written texts; even her adversaries in this discussion at times use this point to try to undermine her argument. It seems important to me theoretically, however, that if we claim that relativity works best when unconscious and habitual, that we be able to modify our thought processes and practices when we bring things to conscious attention. This is in accord with some contemporary research, for example that of Boroditsky (2001), who finds that Mandarin and English speakers use differing predominant metaphors to talk about time (Mandarin speakers use vertical metaphors; English speakers use horizontal). However, after a small amount of linguistic training, English speakers mastered the vertical metaphors as well. This might be an explanation of how gender pronouns can influence people’s performance on standardized test questions in an unconscious way, *and also in* a conscious way.

In addition, it’s important to question whether the relativity axiom applies in artificial language settings. Sapir wrote that “It is quite an illusion to imagine that one adjusts to reality without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection” (*Status of linguistics*

as a science, 187). But couldn't it be argued that, in the context of standardized testing, we are using language to solve specific problems of communication? Standardized test questions are decidedly *not* naturally occurring speech. In my experience writing practice test questions and tutoring, Lily is at least in part correct that test writers are writing questions in a formulaic, patterned way, and many test takers learn how to approach the questions in a such a way as well. So is this a hyper-example of the relativity phenomenon, that test-takers are *even less aware* of grammatical categories *because* they are answering in habitual and "programmed" ways? Or should we believe Sami, on the other hand, who claims that test-takers *are aware* of at least the gendered and racialized names in test questions? The "truth" is, in a sense, irrelevant here, as both these ways of thinking are language ideologies. As such, I argue that they have the potential to mediate between the unequal social structures both Lily and Sami recognize and test-takers' accuracy.

Concluding remarks

While listening to what people say about gender pronouns on a standardized test cannot in the end answer the question of why men do best on masculine pronoun questions and women do best on feminine and neutral pronoun/referent questions, it can help us understand what individuals think is going on. Thus, it provides us access to at least some of the language ideologies that can be articulated in the minds of the test-takers. These language ideologies are multiple, overlapping, and even contradictory, but they are deeply and emotionally felt:

- language makes a difference when we notice it

- standardized tests constitute a special type of language use
- language use (i.e., gendered pronouns and representations) matters in contexts other than standardized tests
- some people are more likely to notice gendered or racialized language than others

I claim that these language ideologies can go a long way towards explaining the meanings and values people associate with language. In order to claim some form of linguistic relativity – that language, thought, and culture are interrelated and co-constitutive – it is not necessary to come up with a theory that only structural (grammatical) or only functional factors are at play. As Lucy writes, “various ideologies of language may play a pivotal role in the essentialization and regimentation of both structure and use” (64). What people *believe* to be happening on standardized test questions may be as important in structuring and influencing what *happens* as anything else.

References

- Boroditsky, Lera. 2001. "Does language shape thought?: Mandarin and English speakers' conceptions of time." *Cognitive Psychology* 43:1-22.
- College Board. 2003. "SAT® Verbal and Math Scores Up Significantly as a Record-breaking Number of Students Take the Test." Accessed electronically on 12/14/06 at <http://www.collegeboard.com/press/releases/26858.html>.
- Guyatt, G. H., D. J. Cook, L. Griffith, S. D. Walter, C. Risdon, and J. Liutkus. 1997. "Attitudes toward the use of gender inclusive language among residency trainees." *Canadian Medical Association Journal* 156:1289-1293.
- Hill and Mannheim. 1992. "Language and World View." *Annual Review of Anthropology* 21:381-406.
- Kroskrity, Paul V. 2000. "Regimenting Languages: language ideological perspectives." In Paul V. Kroskrity (ed.), *Regimes of Language: Ideologies, Politics, and Identities*. Santa Fe, NM: School of American Research Press, 1-34.
- Lucy, John A. 1996. "The scope of linguistic relativity: an analysis and review of empirical research." In J.J. Gumperz and S.C. Levinson (eds.), *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press, 37-69.
- Maltz, Daniel N. and Ruth A. Borker. 1982. "A cultural approach to male-female miscommunication." In John J. Gumperz (ed.), *Language and Social Identity*, Cambridge: University Press, 196-216.
- Ortner, Sherry. 1989. *High Religion: A cultural and political history of Sherpa Buddhism*. Princeton, NJ: Princeton University Press.
- Parks, J. B., and M. A. Robertson. 2004. "Attitudes toward women mediate the gender effect on attitudes toward sexist language." *Psychology of Women Quarterly* 28:233-239.
- Romaine, Suzanne. 1999. *Communicating Gender*. Mahway, NJ: Lawrence Erlbaum Associates.
- Sapir, Edward. 1985[1929]. "The status of linguistics as a science." In D.G. Mandelbaum (ed.), *Selected Writings of Edward Sapir in Language, Culture, and Personality*. Berkeley: University of California Press, 160-166.
- Schneider, J.W., and S.L. Hacker. 1973. "Sex role imagery and use of the generic "man" in introductory texts: A case in the sociology of sociology. , 8, 12-18)." *American Sociologist* 8:12-18.
- Silveira, J. 1980. "Generic Masculine Words and Thinking." *Women's Studies International Quarterly* 3:165-178.
- Walsh, Margaret; Crystal Hickey; and Jim Duffy. 1999. "Influence of item content and stereotype situation on gender differences in mathematical problem solving." *Sex Roles* 41:3/4, 219-240.