He and She: What's the Real Difference?

Clive Thompson

Clive Thompson was born in 1968 in Toronto, Canada, and received a B.A. in political science and English from the University of Toronto in 1987. He began his career writing about politics, but due to his lifelong interest in computers, switched to writing primarily about science technology. When asked to submit his biography for this book, Thompson wrote the following of this piece: "What interested me about this story was how the scientists used artificial intelligence to examine questions about male and female identity that are as old as the hills. Human philosophy and linguistics has for millennia been limited by the fact that human brains are only good at observing small collections of text at a time; when we try to think about the way language works, we rely on our knowledge of the thousands of books and articles we've read in our lifetime. But computers are able to scan millions and billions of pieces of human writing — allowing them to observe patterns that we ourselves would never be able to spot."

In 2002, Thompson was a Knight Science Journalism Fellow at M.I.T. His writing and research is archived online at www.collisiondetection.net. Thompson writes regularly for the New York Times Magazine, Discover, Wired, Details, *and the* Boston Globe, *where this article originally appeared on July 6, 2003. He currently lives in New York.*

Writing to Discover: *Think about what we can learn about the author of a piece of writing aside from what the author tells us directly. Can you tell what a writer is like as a person, a writer's age, or if the writer is a male or female from the style of the writing? Explain how you came to your conclusions.*

Imagine, for a second, that no [author's name] is attached to this article. Judging by the words alone, can you figure out if I am a man or a woman?

Moshe Koppel can. This summer, a group of computer scientists — including Koppel, a professor at Israeli's Bar-Ilan University — are publishing two papers in which they describe the successful results of a gender-detection experiment. The scholars have developed a computer algorithm that can examine an anonymous text and determine, with accuracy rates of better than 80 percent, whether the author is male or female. For centuries, linguists and cultural pundits have argued heatedly about whether men and women communicate differently. But Koppel's group is the first to create an actual prediction machine.

A rather controversial one, too. When the group submitted its first paper to the prestigious journal *Proceedings of the National Academy of Sciences*, the referees rejected it "on ideological grounds," Koppel maintains. "They said, 'Hey, what do you mean? You're trying to make some claim about men and women being different, and we don't know if that's true. That's just the kind of thing that people are saying in order to oppress women!' And I said, 'Hey — I'm just reporting the numbers.' "

When they submitted their papers to other journals, the group made a significant tweak. One of the co-authors, Anat Shimoni, added her middle name "Rachel" to her byline, to make sure reviewers knew one member of the group was female. (The third scientist is a man, Shlomo Argamon.) The papers were accepted by the journals *Literary* and *Linguistic Computing and Text*, and are appearing over the next few months. Koppel says they haven't faced any further accusations of antifeminism.

5 The odd thing is that the language differences the researchers discovered would seem, at first blush, to be rather benign. They pertain not to complex, "important" words, but to the seemingly quotidian parts of speech: the ifs, ands, and buts.

For example, Koppel's group found that the single biggest difference is that women are far more likely than men to use personal pronouns — "I," "you," "she," "myself," or "yourself" and the like. Men, in contrast, are more likely to use determiners — "a," "the," "that," and "these" — as well as cardinal numbers and quantifiers like "more" or "some." As one of the papers published by Koppel's group notes, men are also more likely to use "post-head noun modification with an *of* phrase" — phrases like "garden of roses."

It seems surreal, even spooky, that such seemingly throwaway words would be so revealing of our identity. But text-analysis experts have long relied on these little parts of speech. When you or I write a text, we pay close attention to how we use the main topic-specific words — such as, in this article, the words "computer" and "program" and "gender." But we don't pay much attention to how we employ basic parts of speech, which means we're far more likely to use them in unconscious but revealing patterns. Years ago, Donald Foster, a professor of English at Vassar College, unmasked Joe Klein as the author of the anonymous book *Primary Colors*, partly by paying attention to words like "the" and "and," and to quirks in the use of punctuation. "They're like fingerprints," says Foster.

To divine these subtle patterns, Koppel's team crunched 604 texts taken from the British National Corpus, a collection of 4,124 documents assembled by academics to help study modern language use. Half of the chosen texts were written by men and half by women; they ranged from novels such as Julian Barnes's *Talking It Over* to works of nonfiction (including even some pop ephemera, such as an instant-biography of the singer Kylie Minogue). The scientists removed all the topic-specific words, leaving the non-topic-specific ones behind.

Then they fed the remaining text into an artificial-intelligence sorting algorithm and programmed it to look for elements that were relatively unique to the women's set and the men's set. "The more frequently a word got used in one set, the more weight it got. If the word 'you' got used in the female set very often and not in the male set, you give it a stronger female weighting," Koppel explains.

10 When the dust settled, the researchers wound up zeroing in on barely 50 features that had the most "weight," either male or female. Not a big group, but one with ferocious predictive power: When the scientists ran their test on new documents culled from the British National Corpus, they could predict the gender of the author with over 80 percent accuracy.

It may be unnerving to think that your gender is so obvious, and so dominates your behavior, that others can discover it by doing a simple word-count. But Koppel says the results actually make a sort of intuitive sense. As he points out, if women use personal pronouns more than men, it may be because of the old sociological saw: Women talk about people, men talk about things. Many scholars of gender and language have argued this for years.

"It's not too surprising," agrees Deborah Tannen, a linguist and author of best-sellers such as *You Just Don't Understand: Women and Men in Conversation.* "Because what are [personal] pronouns? They're talking about people. And we know that women write more about people." Also, she notes, women typically write in an "involved" style, trying to forge a more intimate connection with the reader, which leads to even heavier pronoun use. Meanwhile, if men are writing more frequently about things, that would explain why they're prone to using quantity words like "some" or "many." These differences are significant enough that even when Koppel's team analyzed scientific papers — which would seem to be as content-neutral as you can get — they could still spot male and female authors. "It blew my mind," he says.

But this gender-spotting eventually runs into a $64,000 conceptual question: What the heck is gender, anyway? At a basic level, Koppel's group assumes that there are only two different states — you're either male or female. ("Computer scientists love a binary problem," as Koppel jokes.) But some theorists of gender, such as Berkeley's Judith Butler, have argued that this is a false duality. Gender isn't simply innate or biological, the argument goes; it's as much about how you act as what you are.

Tannen once had a group of students analyze articles from men's and women's magazines, trying to see if they could guess which articles had appeared in which class of publication. It wasn't hard. In men's magazines, the sentences were always shorter, and the sentences in women's magazines had more "feeling verbs," which would seem to bolster Koppel's findings. But here's the catch: The actual identity of the author didn't matter. When women wrote for men's magazines, they wrote in the "male" style. "It clearly was performance," Tannen notes. "It didn't matter whether the author was male or female. What mattered was whether the intended audience was male or female."

15 Critics charge that experiments in gender-prediction don't discover inalienable male/female differences; rather, they help to create and exaggerate such differences. "You find what you're looking for. And that leads to this sneaking suspicion that it's all hardwired, instead of cultural," argues Janet Bing, a linguist at Old Dominion University in Norfolk, Virginia. She adds: "This whole rush to categorization usually works against women." Bing further notes that gays, lesbians, or transgendered people don't fit neatly into simple social definitions of male or female gender. Would Koppel's algorithm work as well if it analyzed a collection of books written mainly by them?

Koppel enthusiastically agrees it's an interesting question — but "we haven't run that experiment, so we don't know." In the end, he's hoping his group's data will keep critics at bay. "I'm just reporting the numbers," he adds, "but you can't be careful enough."

FOCUSING ON CONTENT

1. Describe the gender-detection experiment performed by computer scientists at Israel's Bar-Ilan University. How was the experiment set up and carried out?
2. What were the results of the experiment?
3. What were the original concerns of the editors of *Proceedings of the National Academy of Sciences* when the researchers submitted the results of their experiment? How did the researchers respond to the concerns other editors had?
4. Why doesn't Deborah Tannen find the results of the research that Koppel and his associates did surprising?
5. What question(s) are not answered by Koppel's research, according to linguist Janet Bing?

FOCUSING ON WRITING

1. What words are women far more likely to use? What words are men more likely to use? Why are the words in both cases rather surprising?
2. What is a "post-head noun modification with an *of* phrase"(6)? What is the example that Thompson gives? Are men or women more likely to use the construction?
3. Review paragraph 14 and explain how the research that Tannen did with her students extends the findings of the research that Koppel and his associates did. What role does audience play in the kinds of language that writers use? (Glossary: *Audience*)
4. Why do you suppose Thompson ends his article with a reiteration of the "I'm just reporting the numbers" quotation that he used earlier in his article? To what does Koppel refer when he's quoted at the end of the article by saying, "but you can't be careful enough"?

LANGUAGE IN ACTION

Using the tips that Clive Thompson says are at the heart of the new program developed to detect whether an author is likely male or female as well as the indicators provided in Nathan Cobb's article (pp. 300–04), examine the following passages to see if you can make a calculated guess as to the sex of their authors. Make sure you are able to explain to your instructor or the members of your class why you could or could not make a judgment in each case. (The authors' names are found on p. 348)

WRITER 1

I was saved from sin when I was going on thirteen. But not really saved. It happened like this. There was a big revival at my Auntie Reed's church. Every night for weeks there had been much preaching, singing, praying, and shouting, and some very hardened sinners had been brought to Christ, and the membership of the church had grown by leaps and bounds. Then just before the revival ended, they held a special meeting for children, "to bring the young lambs to the fold." My aunt spoke of it for days ahead. That night I was escorted to the front row and placed on the mourners' bench with all the other young sinners, who had not yet been brought to Jesus.

My aunt told me that when you were saved you saw a light, and something happened to you inside! And Jesus came into your life! And God was with you from then on! She said you could see and hear and feel Jesus in your soul. I believed her.

WRITER 2

The stealth of autumn catches one unaware. Was that a goldfinch perching in the early September woods, or just the first turning leaf? A red-winged blackbird or a sugar maple closing up shop for the winter? Keen-eyed as leopards, we stand still and squint hard, looking for signs of movement. Early-morning frost sits heavily on the grass, and turns barbed wire into a string of stars. On a distant hill, a small square of yellow appears to be a lighted stage. At last the truth dawns on us: Fall is staggering in, right on schedule, with its baggage of chilly nights, macabre holidays, and spectacular, heart-stoppingly beautiful leaves. Soon the leaves will start cringing on the trees, and roll up in clenched fists before they actually fall off. Dry seedpods will rattle like tiny gourds. But first there will be weeks of gushing color so bright, so pastel, so confettilike, that people will travel up and down the East Coast just to stare at it — a whole season of leaves.

WRITING SUGGESTIONS

1. In paragraph 13, Clive Thompson writes of the research that Koppel's group has done: "But this gender-spotting eventually runs into a $64,000 conceptual question: What the heck is gender, anyway? At a basic level, Koppel's group assumes that there are only two different states — you're either male or female. ('Computer scientists love a binary problem,' as Koppel jokes.)

But some theorists of gender, such as Berkeley's Judith Butler, have argued that this is a false duality. Gender isn't simply innate or biological, the argument goes; it's as much how you act as what you are." Write an essay in which you attempt to define the term *gender* using Thompson's essay as well as other sources that you find in your library or on the Internet.

2. If Deborah Tannen is correct, that the most important issue in word choice is the writer's intended audience, then it would seem that audience as a writer's concern is perhaps even more important than we have assumed. We are never sure who will read what we write, but we need an audience in mind as we write. Or do we? Is it possible to write for ourselves or for an audience so general that we don't have it clearly in mind? Write an essay in which you examine the concept of audience as it pertains to the writer's craft. Is it as important as writing teachers and theorists think? If so, why? What have writing experts said about audience that is important for us to know?