Why Your Computer Doesn’t Understand You

James Pustejovsky has his eyes on the prize—a computer that talks in any language and knows what’s in your fridge.

ENCORE!

In creating music, these award-winning composers know the score—anything goes as long as it’s good.
Dear Reader,

Catalyst, the Brandeis University research magazine for the sciences, social sciences, and humanities, chronicles activities that make this campus an exceptional center of learning, teaching, and scholarship. From new programs focusing on entrepreneurship and computational linguistics, to forefront research on race and free speech in the nineteenth century and the biology of memory, Brandeis is helping define liberal arts for the twenty-first century. Catalyst is one compelling way we are telling that story.

In this issue, James Pustejovsky, professor of computational linguistics, describes the complex challenges that artificial intelligence faces in communicating through language—any language—the right nuances of meaning and word senses. You can try your hand at computational linguistics by taking the quiz on page twelve.

The language of music—composition—is also thriving at Brandeis. In these pages you will discover how three award-winning composers are creating dynamic new music that embraces diverse sounds and styles. From the early days when Aaron Copland and Leonard Bernstein were faculty members, music composition has been a particular forte here.

You’ll also learn how psychologist Derek Isaacowitz has pioneered the use of eye-tracking technology to show that, contrary to conventional wisdom, older adults do not view the world negatively, but rather are able to maintain an even, sunny outlook. But it’s a different story for young people.

I am confident you will enjoy these stories and the other profiles, commentary, and news briefs in Catalyst.

Sincerely,

Adam B. Jaffe
Fred C. Hecht Professor in Economics and Dean of Arts and Sciences
## 2 THE PIPELINE

Brandeis is nurturing a culture of entrepreneurship through expanded industry relationships and new academic programs that incorporate business-building know-how.

## 3 NOTABLE RESULTS

A towering figure in labor law, Jules Bernstein '57 made his name fighting for workers’ rights.

## 4 ENCORE!

In creating music, Brandeis’s award-winning composers know the score—anything goes as long as it’s good.

## 8 WHY YOUR COMPUTER DOESN’T UNDERSTAND YOU

James Pustejovsky has his eyes on the prize—a computer that talks in any language and knows what’s in your fridge.

## 13 HANDS ON

Taisha Sturdivant ‘11 has never let her tragic background get in the way of life or her plans to improve it for others.

## 14 DEREK ISAACOWITZ IS IN THE MOOD

The psychologist’s eye-tracking research reveals rosy results in older people.

## 16 OBSERVATIONS

Soviet area studies helped us understand our cold-war enemy. So why not follow the same model to understand the terrorist threat or Afghanistan?

## 18 RESEARCH NOTES

From love and death in lyric poetry, to research on high-energy physics and active matter, Brandeis scholars generate new knowledge to make the world a better place.

## 20 THE SHORT, SIGNIFICANT LIFE OF THE FRIEDLAND BIOCHEMISTRY BUILDING

How one building came to epitomize the pace of scientific discovery at Brandeis.
James Pustejovsky has his eyes on the prize—a computer that talks in any language and knows what’s in your fridge.

By Theresa Pease
Therapist: Hello, I am Eliza. How can I help you?

Patient: I need my head shrunk really fast.

Therapist: Why did you say your head shrunk really fast?

Patient: I didn’t say it shrunk. I said I need you to shrink it.

Therapist: You are being a bit negative.

Patient: I’m going to jump off a bridge. I guess you could call that negative.

Therapist: Do you wish to guess me?

Patient: No, I wish to swing from a tree screaming nasty words.

Therapist: You are being a bit negative.

Patient: Just as I thought; you can’t help me. Life stinks. Don’t you think life stinks?

Therapist: Oh, I think life stinks.

Believe it or not, dialogues like this one featuring MIT computer scientist Joseph Weizenbaum’s program Eliza riveted the nascent artificial intelligence community forty-five short years ago. Designed to simulate a Rogerian psychotherapist, Eliza parroted back key phrases so convincingly that some users scarcely believed they were speaking to a machine. Despite the awkward phrasing and frequent errors, philosophers, linguists, and computer scientists the world over thought that the day when you could comfortably converse with a robotic assistant was just around the corner.

Fast-forward almost half a century and you’ll find another generation of scholars far more cautious in their estimate of exactly when you’ll be able to trade gossip with R2-D2, C-3PO, or, for that matter, your automobile. Just think about the chats that take place when you are, say, on hold with American Airlines, and you’ll realize that Eliza, history’s most famous therapist wannabe, was but a chimera.

EASILY CONFUSED CYBERSOURCES

It’s not that progress has not been made. If you want to know yesterday’s sports scores, learn whether your neighbor’s house is for sale, or get the names of Snow White’s seven dwarves, you’ll surely be able to find your answer with a few clicks of the mouse. You can get reasonably accurate directions from your TomTom and dispute consumer problems with a robotic customer-service agent on eBay. But try to ferret out more open-ended data—say, what books have been written about post–Civil War reconstruction since 1942—and you’ll confuse your cybersources woefully.

That, says James Pustejovsky, head of the linguistics section in Brandeis’s Department of Computer Science, is because Google and its cousins have no way to discern references to the U.S. Civil War from any other sentences that contain the words “civil” and “war,” and no knowledge of the word “since.” In fact, the search engine has no sense of time at all. To make matters worse, computers have no appreciation for irony, and they don’t understand the relationship of words within a sentence: whether Joe killed Mary or Mary killed Joe is all the same to them. They are not good at synonyms; if you ask, “Who purchased the Sears Tower?” your search engine may completely skip a headline about someone who “bought” or “acquired” the building. Further, such programs do not grasp common idioms or have a command of what Pustejovsky calls “the social clues.” Greet Eliza with “Give me five, Baby,” and she’ll respond with, “Don’t you ever say hello?”

JAMES PUSTEJOVSKY
Greet Eliza with “Give me five, Baby,” and she’ll respond with, “Don’t you ever say hello?”

trillion numbers and it will make sense out of them. But if you tell a computer, ‘When the alarm rang, Johnny got up from his desk, went out the door, and left the house,’ it won’t be able to follow the story. It can’t figure out what happened, why, or in what order.”

Indeed, it was an underestimate of how much knowledge the human brain really contains that caused mid-twentieth-century artificial intelligence pioneers to misjudge the enormity of the task before them. Discovering how the brain was “hardwired” and then mimicking those processes to produce machine “intelligence” sounded infinitely more doable to Weizenbaum’s colleagues than it does to contemporary researchers.

But the difficulty doesn’t dilute the thrill of the quest, says Pustejovsky, who for the past twenty years has led a Brandeis research group trying to teach a computer to visualize and re-create a simple scenario like Johnny’s.

THE DARK MATTER OF SEMANTICS

“If you understand English,” Pustejovsky notes, “you can create a spatial and temporal narrative about Johnny. You understand the implied context: The desk was in his house, there was a chair at the desk, and to get to the door Johnny changed his posture and walked. He probably turned a doorknob. Understanding the narrative’s hidden meaning is a large part of artificial intelligence, and all our efforts are aimed at building an algorithm that will make it possible for machines to do that.”

Understanding this complex subtext (Pustejovsky calls it the “dark matter” of semantics) involves multiple parameters in the brain that he and his team are striving to replicate in the computer. Noting that such efforts involve studying the way children learn to tell stories, Pustejovsky says, “I’m not a cognitive psychologist, but I am really interested in cognitive development, because these computational models will succeed only insofar as they reflect the way we ourselves learn.”

Mathematically inclined as a youngster, Pustejovsky first became curious about language when as a high-school student he went on a scuba-diving trip in Mexico. During a day ashore, he visited Mayan ruins and was riveted by the hieroglyphs he saw.

Although the youth entered MIT with the intention of majoring in electrical engineering or math, he heard in his sophomore year about a fascinating syntactical analysis class his roommate was taking with world-renowned linguist Kenneth Hale. It was the first year of a new undergraduate degree program in linguistics and philosophy, and Hale shared faculty billing with other superstars like the linguist Noam Chomsky and John Robert Ross, a syntactician.

“That day,” Pustejovsky says, “I dropped physics and signed on for Ken’s class. Within two weeks, I had dropped two more courses and was enrolled in three linguistics classes.”

Pustejovsky was no polyglot. In fact, he grew up in a bilingual English- and Czech-speaking home in Texas without amassing a Czech vocabulary of more than four dozen words. Instead, it was the technical aspects of language that fascinated him.

“Linguistics was analytic, and it was a lot like math,” he says. “It was pattern analysis, really, looking for the structure of language. I wasn’t interested in the emotive content, and I certainly was not drawn in by the literary aspect; I had never been excited by English classes. It was the scaffolding, the underlying architecture of language, that got my attention.”

APPLYING MATHEMATICAL PRINCIPLES TO LANGUAGE

By making logical and algebraic connections, Pustejovsky realized, he could apply mathematical principles to language. This enabled him to mine various tongues for similarities and differences in word order, sentence construction, subject-verb agreement, and other properties of speech that shed scientific light on the brain’s innate ability to receive and send communications. It was like solving an arcane and very complex puzzle.

After graduating from MIT in 1978, Pustejovsky received a DAAD grant—a German Fulbright Scholarship—to explore linguistics and philosophy at Philipps University in Marburg, Germany. In 1985, he received a doctorate in linguistics from the University of Massachusetts at Amherst, then stayed around for an eighteen-month postdoc in artificial intelligence. After joining the Brandeis computer science faculty in 1986, he continued to do linguistics research, often with funding from the National Science Foundation, National Institutes of Health, or DARPA (Defense Advanced Research Projects Agency), and to publish theoretical linguistics articles in scholarly journals.

Linguistics classes, however, took place in Brandeis’s psychology department until 2005, when the undergraduate
linguistics program was reorganized as an interdepartmental program. Pustejovsky joined the program faculty and now serves as chair. The program is designed to provide a broad, liberal-arts view of linguistics, while the new master’s program concentrates more heavily on understanding how languages operate in order to put that information to work in artificial intelligence applications.

FROM ENGLISH TO MANDARIN
“We chose to move the Brandeis program in this direction because computational linguistics is a growth industry. People need to develop communication and language software for the Web, for industries, for government, and for translation purposes. This is not about creating applications that make money for Google. This is about offering a solid, cross-disciplinary, foundational program that has ties to anthropology, classics, psychology, and neuroscience,” says Pustejovsky, recently named to the TJX/ Feldberg Chair in Computer Science.

The fledgling graduate program has two tracks: a five-year BA/MA degree and a two-year master’s program. The first class of five-year students graduated in spring 2009, and the inaugural group of two-year MA candidates arrived last fall.

Among them are students who aspire to teach at the university level, develop applications for various industries, and put computational linguistics skills to work deciphering ancient scripts. Other potential applications of computational linguistics include facilitating literary analysis and the study of literary history; performing complex archival tasks; and enabling the cataloging and retrieval of photographs and other images that do not include text, Pustejovsky notes.

“The Holy Grail of computational linguistics and artificial intelligence,” he says, “would be a reliable hand translator. Someone could speak English into a device in Beijing, and out would come a translation in Mandarin, with just the right word senses and nuances of meaning.

“If we are successful in our efforts,” Pustejovsky says, “more mundane aspects of our lives will also be impacted by this work. For example, the time will come when you will be able to ask your refrigerator what you need to buy at the market, and it will tell you, ‘Well, the milk has been in here for two weeks, and the fruit section is empty.’ Or maybe it will just print out your shopping list for you.”

Theresa Pease is senior editor in the Office of Communications and editor of Brandeis University Magazine.

BREAK THE CODE

TENJI KARAOKE
Braille is a tactile writing system, based on a series of raised dots, that is widely used by the blind. It was invented in 1821 by Louis Braille to write French but since has been adapted to many other languages. English, which uses the Roman alphabet just as French does, required very little adaptation, but languages that do not use the Roman alphabet, such as Japanese, Korean, or Chinese, are often organized in a very different manner.

Below is a Japanese word written in the tenji (“dot characters”) writing system. The large dots represent the raised bumps; the tiny dots represent empty positions.

karaoke  •  •  •  •  •  •  •  •  •

The following tenji words represent atari, haiku, katana, kimono, koi, and sake. Which is which? You don’t need to know either Japanese or Braille to figure it out; you’ll find that the system is highly logical.

A.  •  •  •  •  •  •  •  •  •
B.  •  •  •  •  •  •  •  •  •
C.  •  •  •  •  •  •  •  •  •
D.  •  •  •  •  •  •  •  •  •
E.  •  •  •  •  •  •  •  •  •
F.  •  •  •  •  •  •  •  •  •

What are the following words?

G.  •  •  •  •
H.  •  •  •  •

Answers: A. haiku  B. sake  C. katana  D. kimono  E. koi  F. atari  G. anime  H. anime