

Logically speaking

An associate professor and an inquisitive high school student discover that a computer can learn language as quickly as a child.

BY DICK LLOYD

Ann Kromsky rattles off the names like they were old friends — noted researchers in psychology like Chomsky, Elman, Saffran, Marcus, Leninburg and Pinker. She cites their work with the vigor of a veteran researcher as she describes her own studies at UCR into how children learn to speak.

In fact, Kromsky, just 17 and not quite graduated from Corona High School, is already challenging the notion of some of these experts that the learning of language is governed by innate biological factors.

In research conducted in collaboration with UCR Associate Professor of Psychology Curt Burgess and Computational Lab graduate student Kevin Lund, Kromsky has demonstrated that a computer can learn language just as quickly as a child can. The results of her studies strongly suggest that children learn language simply by hearing it. As a practical matter, the research has profound implications for the perfection of artificial intelligence and other applications as well.

"It is the first time this has ever been done, using a computer to derive (word) meaning without humans telling it what to do," said Kromsky.

Kromsky's study, "A Computational Model of Child Memory: Memory Acquisition Without Biological Constraints," is so impressive it won her a \$10,000 scholarship in the prestigious Westinghouse Science Talent Search finals in March in Washington, D.C.

Last year this same original research, which she has been conducting under mentor Burgess since she was a 15-year-old high school sophomore, won first place and a \$4,000 scholarship at the California Junior Science and Humanities Symposium sponsored by the U.S. Army and Navy. Kromsky is the first female to win the annual Junior Science contest.

In the Westinghouse competition, Kromsky was seventh in the nation out of an original field of 1,582 entrants.

Importantly, she accomplished all this just four years out of the Soviet Union, coming to the U.S. from Baku, Azerbaijan as a 13-year-old speaking practically no English. Today, the honors student is thoroughly fluent in English with a formidable vocabulary. She'll graduate in June, ranked fourth in her class of 547.

Actually, it was her 1995 historical research into the

origins of the U.S. purchase of Alaska that led to her interest in language research and contact with Burgess.

Remembering friends in her native land who believe that Alaska rightfully belongs to Russia, Kromsky decided to research the truth for herself. She delved into the U.S. National Archives and other sources and concluded that indeed Alaska was a legitimate purchase, signed off by President Andrew Johnson and Czar Alexander II of Russia, even if it only cost a mere two cents an acre

That first foray into research, in which she entered the National History Day Contest and won the California competition and then went on to the national finals in Washington, D.C., not only proved her skill as a researcher but whetted her appetite for more. The full experience and contact with other high school researchers made her realize she would like to engage in science research and participate in the Junior Science competition, which required working with a university professor.

Because of her interest in psychology and language, she was led to Burgess, who admits to his original skepticism at working with a high school student on such complex issues as language learning and computer modeling.

"I was not expecting a high school student to understand this work," Burgess remembers. "I couldn't have been more wrong."



Associate Professor of Psychology Curt Burgess admitted to initial skepticism at working with high school student Ann Kromsky. "I couldn't have been more wrong," he said.

Not only did she thoroughly absorb materials Burgess asked her to study as a test of her interest and skill, she returned with profound questions and suggestions. And, rather than study possible differences in the way males and females learn language as Burgess originally suggested, Kromsky proposed to study how children learn language and suggested ways to go about it.

Burgess agreed.

Until now, the research literature, which Kromsky simply devoured (and questioned), has widely suggested that because language is so chaotic and noisy and irregular a child must have some innate mechanism facilitating the learning of speech. Otherwise, it is reasoned, the child could make no sense of the myriad complex speech particles bombarding the senses.

continued p. 10

Kromsky suggested that creation of a computerized model of the speech of children could test that hypothesis. By loading the computer model with words and utterances of children, she wondered, could the computer in fact learn new language just as children learn language? If it were true that the learning of language is somehow innate or biologically controlled, this would handicap a computer model to learn the meaning of a word or to learn new language, Kromsky reasoned. A computer, after all, has no biological predisposition.

Kromsky created a computer model based on Burgess and Lund's powerful adult language and memory computer model — Hyperspace Analog to Language (HAL) — to test her ideas. But before the new model, dubbed HALjr. could be used, she had to create a sufficiently large database of childhood words or utterances. In addition, it seemed there were not enough sources to collect the volumes of child speech or conversational language that would be needed.

Perhaps transcribing from child television networks like Nickelodeon could be a solution, or maybe the Internet, possibly tapping into children's conversations in chat rooms. Yet, those options would have been terribly ponderous and time consuming. But onto the Internet Kromsky strode, searching for that golden source. There she found it, a treasury of conversational speech between parents and children of different backgrounds, race and economic conditions—millions of words collected over many years at Carnegie Mellon University in Pittsburgh, Pennsylvania.

So, taking a specialized computer class in data manipulation, Kromsky separated out the children's utterances and

produced a rich field of 1.7 million words to plug into her model.

But it was not simply words, it was what the computer did with those words.

"It (HALjr.) formed a high dimensional space, like a memory map, where each word had its own vector, its own place," said Kromsky. From her huge bank of words children use, the computer concentrated on words that occurred at least 50 times. These were used to create word vectors, or history of usage. Words, usage and location within sentences were translated into a high-dimensional memory matrix by which the computer analyzed word placement in relation to other words, thus eventually deriving meaning.

While it appears complicated, Kromsky says the process is rather simple. In the memory matrix, the computer is able to establish meaning or definition of the words through context—their history of occurrence with other words. "This is a revolutionary feature," says Kromsky, "because others believe that what defines a word are the features of the word, like the features of a dog or cat. What we are saying is that what defines a word is its context."

The computer not only came to define or understand new language but it did as it quickly as a child learns, said Kromsky. In only six to eight contacts with a word, it came to understand its use. In fact, the experiments revealed that acquisition of meaning can occur in as little as one or two trials. Typically, a child learns to use a word in about six to eight encounters.

HALjr. demonstrates that the basic building blocks of language (semantics, grammatical class, semantic-syntactic relationships) can, in many ways, be explicitly modeled with very straightforward

learning mechanisms, and with fairly complex child utterances, said Kromsky.

Kromsky admits that engaging in research as a high school student is unusual, but she affirms rather emphatically that she is a typical teenager. She could have graduated early but chose to stay with her classmates to graduate this June. And while thoroughly engrossed with research, she is equally wrapped up in fun things at high school. She ran for Homecoming Queen, is president of the Corona High School Key Club and devotes lots of effort to the Science Club and the International Studies Club. She loves to sing in the Corona High Madrigals, which are recognized as among the state's very best madrigals groups.

She enjoys living in the U.S., the vast opportunities, the fast pace, the people, even the free-ways.

Frightening changes in her country, including tanks in the streets as it fell into internal conflict and social disintegration following the collapse of the Soviet Union, led her father Alex and mother Marina and younger brother Yuri and Ann to depart and come to the U.S. where they would join relatives in Corona.

Though she has a bright future in the U.S., Kromsky is sure she'll return to Russia and Azerbaijan one day, if only to visit. Right now, it's on to more schooling and more research and perhaps studies to become a physician.

With her scholastic success, it isn't surprising that Kromsky has received a number of scholarship offers from universities across the country. In fact, it's turning out to be a difficult choice for her to make, choosing between the West and the East.

Oh, that we all had such decisions.