

Researchers mine millions of metaphors through computer-based techniques

March 3 2009, By Lisa M. Krieger

Metaphors cannot be taught, asserted the great philosopher Aristotle. "It is the one thing that cannot be learnt from others." But a computer scientist and literary historian say he's wrong.

In a project started at Stanford University, the researchers are teaching computers how to analyze texts from Plato to Pynchon, mining millions of these abstract phrases. (Metaphorically speaking.)

They're building a vast searchable database, making it possible to browse historic patterns of word usage - for instance, "rose" and "love" - from ancient Homeric epics to postmodern cyberpunk novels, and everything in between.

"As a tool, it provides a really powerful way of thinking about a lot of literature at once," said English literature professor Brad Pasanek, who collaborated on the project with longtime friend and computer scientist D. Sculley.

The work makes tangible what the German linguist Harald Weinrich called the "metaphoric field." "Pasanek's database is the first 'metaphoric field' that we can actually see and use," said Franco Moretti, a Stanford comparative literature professor. "It provides empirical proof for a daring, but never wholly solid concept."

This approach to studying literature was inconceivable back around 330 BCE, when Aristotle wrote that "the greatest thing by far is to be a

master of the metaphor," language that compares seemingly unrelated subjects - a "winged thought," for instance.

But two new trends have created a field of computer-based literary analysis, part of the emerging discipline called "digital humanities," an intersection of computing and the study of languages, history, philosophy and religion.

Digitized libraries have put an ocean of books - including obscure ones - at readers' fingertips. Using new data mining techniques and "machine learning," researchers can search the millions of words contained in those books to study subtle shifts in how words were used. Analyzing such patterns offers insights to how language - and culture - evolved.

The idea was conceived when Pasanek was idly flipping through his worn copy of "Pride and Prejudice," its key phrases highlighted in bright colors.

Through the tangled tale of Elizabeth, Darcy and Wickham, "marking words that occurred again and again, I realized that you could flip through a novel and see these motifs appear in an explosion of color, then disappear."

The computer replaces the colored marker, he said. "It's possible to trace when and where something appears, what it means, and how it changes," he said.

Pasanek's near-obsessive collection of interesting metaphors began while he was at Stanford working on his Ph.D. First he kept a list on the back pages of the works of Shakespeare, Milton and the King James Bible. As his list grew, he moved to index cards.

"Metaphors are a fundamental figure of speech," he said. "They show

how we think, and how what we think changes over time."

Recognizing he needed help, Stanford computer scientist Matt Jockers helped him create a digital database, which was initially posted in 2005. The list quickly grew to 1,000, then 3,000 entries. But the list's expansion created a special search challenge.

"The nature of metaphor is such that it does not lend itself to easy detection by the usual sorts of pattern matching algorithms," Jockers said. Finding a simile is a fairly straightforward task: one writes a program that looks for text strings of the type "like" and "as."

"Structurally speaking, the phrase 'my love is a red rose' is very much the same as 'my dog is a blue heeler,'" Jockers said. "The former is metaphor, the latter is not."

Pasanek provided the computer with examples of metaphors and "trained" the machine to recognize them. They programmed "proximity searches" between words likely to be metaphoric. For example, a search for "mind" within 100 characters of "mint" finds the following couplet in William Cowper's poetry: "The mind and conduct mutually imprint /And stamp their image in each other's mint."

A similar technique, said Sculley, is used in spam-recognition software.

In one project, they tracked the evolving references to the young mind. In the fourth century B.C., it was referred to as a "tabula rasa," Latin for "blank slate." By the 17th century, John Locke called it a "white Paper, void of all Characters." In 18th century texts, it was compared to a "roasting jack," conjuring up an image of meat spinning on a rotisserie, cooked by flames. As tools changed -- slates, paper, rotisseries -- so did the references.

There are other metaphor databases, though Pasanek says his is the largest and geared toward the history of thought. However, the database (mind.textdriven.com) is still in its beta version, said Pasanek, who now teaches literature at the University of Virginia in Charlottesville. Under renovation, it suffers from what he calls "bug plagues." But with time, it will improve, and broaden to vast horizons.

"A metaphor has a career, and it tells a complete story," he said, "about how we think about ourselves and the world."

*(c) 2009, San Jose Mercury News (San Jose, Calif.).
Visit MercuryNews.com, the World Wide Web site of the Mercury News,
at www.mercurynews.com
Distributed by McClatchy-Tribune Information Services.*

Citation: Researchers mine millions of metaphors through computer-based techniques (2009, March 3) retrieved 29 January 2023 from <https://phys.org/news/2009-03-millions-metaphors-computer-based-techniques.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--