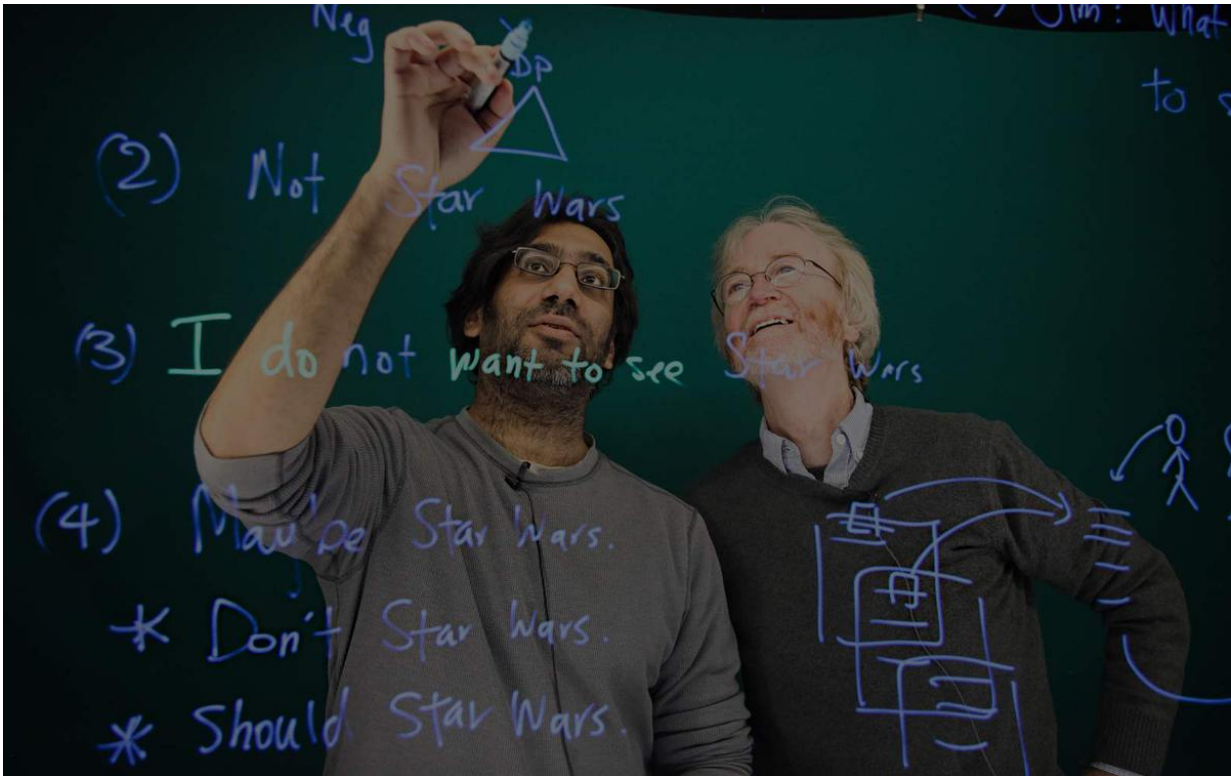


# The linguistic theory of fragment structures

June 15 2016, by Scott Rappaport

---



Credit: Carolyn Lagatutta

How do you say what you mean—without saying what you mean?

That question is more crucial to technological communication these days than you might imagine—particularly in a world where talking with your smartphone, your television, your car, and your house becomes a more commonplace experience every day.

"A lot of talk is fragments—it's the kind of thing we understand reflexively as human beings, but it's much harder for machines," notes Jim McCloskey, professor of linguistics at UC Santa Cruz. "Linguistic theory teaches us what kind of structures there are in our mind, but how to make sense of these fragments is also a nuanced engineering problem."

This problem is one that appeals to a researcher like McCloskey, who has dedicated his work to understanding language, and now Silicon Valley tech companies that are seeking to make mobile devices—phones, tablets, and more—that can understand and decode the subtleties of human language.

And in the search for solutions, UC Santa Cruz students helping with this research have found they are able to apply their knowledge and research skills after graduating as analytical linguists for tech companies big and small.

Asking your phone questions and receiving the correct information can seem astonishing—until the virtual assistant stumbles and doesn't appear to understand a slightly more complex request.

Jim McCloskey and Pranav Anand, both professors of linguistics at UC Santa Cruz, discuss the engineering problems presented by linguistic theory.

## **Autocorrect, yes; fill in the blanks, no**

McCloskey notes that speakers and writers often leave out informationally redundant grammatical material—such as when the verb "call" is omitted in "Jay Z called, but Beyoncé didn't." This process, known as ellipsis, is widespread across the languages of the world, and is particularly common in informal language and dialogue.

Among the many varieties of ellipsis is "sluicing," where what is omitted is not a verb, but an entire sentence. For example, a speaker may leave out the understood sentence "he called" after "why" in a sentence like: "He called, but I don't know why [he called]."

Ellipsis creates challenging scientific and engineering problems. Although research over the past 50 years has shown that the principles permitting ellipsis involve many different types of information (grammatical structure, context, real-world knowledge), the precise mix of these principles and their interaction is still an open question.

Progress to date has been delayed by the lack of one crucial resource: databases that are large enough to validate theories and rich enough to form the basis for machine learning.

At UC Santa Cruz, McCloskey is collaborating with faculty and students in the language sciences to develop that resource—a richly annotated database of naturally occurring ellipsis, which will be freely available to researchers around the globe who are trying to understand what their implications might be for our understanding of the nature of human language.

## **Strong collaboration with undergraduates**

The project, which began with backing from UC Santa Cruz's Institute for Humanities Research in 2013, is now funded by a three-year grant from the National Science Foundation running through the end of 2018.

UC Santa Cruz linguistics professor Pranav Anand, principal investigator on the grant, noted that the reputation of the campus's undergraduate program in linguistics was a primary reason they received the NSF grant.

"They knew we have this army of sophisticated undergraduates who can

do the work," said Anand. "We're very hands-on and workshop-oriented. We don't use textbooks; instead we say, 'here's a problem, let's collaborate.'"

"Even after a few courses, the students are able to do sophisticated annotations," he added. "They are able and up to the task. We hope to collect 30,000 samples minimum over the three years of the NSF grant."

Although the UC Santa Cruz program is focused on theoretical linguistics, Anand said it is also driven by the needs and curiosities of the undergraduate students, who are learning new relevant skills working on this project.

"We are currently heavily recruiting students for sophisticated annotations," he noted. "There's a pipeline—students get doctorates and are now working in Silicon Valley."

"We first noticed it three years ago," added McCloskey. "We realized with our graduate program that students were not going into academic jobs, but rather to Silicon Valley. Their training in statistics and computational design is what new managers say helped prepare them for the job."

But Anand noted that the UC Santa Cruz Linguistics Department is still theoretically based. The pipeline to Silicon Valley is a fortuitous by-product of shared interests.

"You become an expert," he explained. "(Students) are doing case after case, so they're seeing all the patterns, and they discover new forms of fragments. The students are actually producing and creating new knowledge with their work."

Provided by University of California - Santa Cruz

Citation: The linguistic theory of fragment structures (2016, June 15) retrieved 26 April 2024 from <https://phys.org/news/2016-06-linguistic-theory-fragment.html>

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.