

Linguist uses Internet to study how we say things

January 4 2010, By Linda Glaser

(PhysOrg.com) -- Mats Rooth, a Cornell linguist, will use software to study distinctions of prosody (rhythm, stress and intonation) in language by hunting for word patterns on the Internet.

How would you analyze the contents of a million books? Or a million podcasts? Mats Rooth, Cornell professor of linguistics and computing and information sciences, will do it by using software to search for word patterns in text transcriptions of audio and video files.

Rooth is one of eight winners of an international competition, Digging into Data, that challenged scholars to devise innovative humanities and social science research projects using large-scale data analysis. His project, Harvesting Speech Datasets for Linguistic Research on the Web, is based on a pilot project Rooth conducted with graduate student Jonathan Howell. It will look at distinctions of prosody (rhythm, stress and intonation) in spoken [language](#).

According to Rooth, native speakers easily identify what prosody is appropriate in a given sentence, but hypotheses explaining why people have this ability have been controversial to prove because of the difficulty of identifying enough examples of a given phenomenon. "Many of the things we study are so immediate and yet so subtle," he said.

Using the Internet to harvest hundreds or thousands of examples of spontaneous rather than lab-created use of word patterns will enable

researchers to evaluate theories about the form and meaning of prosody on an unprecedented scale. Rooth expects his project to have a transformative effect on the understanding of prosody.

"I'm very excited," Rooth said. "It's a new methodology, and we think a lot of new information will come out."

Four leading research agencies sponsored the Digging into Data competition, with the intention of encouraging international partnerships: the National Endowment for the Humanities, the National Science Foundation, the United Kingdom's Joint Information Systems Committee, and Canada's Social Sciences and Humanities Research Council. Approximately \$2 million will be divided among the eight winners.

Linguist Michael Wagner of McGill University is Rooth's international partner on the project. The Cornell team will be responsible for data retrieval and programming, while McGill researchers will focus on data analysis.

The computer programs, datasets and research products developed in the project will be openly available to the research community via a Web site, [confluence.cornell.edu/display ... ody/Prosody+Datasets](https://confluence.cornell.edu/display/...ody/Prosody+Datasets) . The Web site already contains a sample dataset which, when played, provides a fascinating cacophony of voices saying "than I did," demonstrating the wide range of meaning arising from varied intonation.

Provided by Cornell University

Citation: Linguist uses Internet to study how we say things (2010, January 4) retrieved 23 April 2024 from <https://phys.org/news/2010-01-linguist-internet.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.