Joshua Tauberer

http://razor.occams.info

Joshua Tauberer is a University of Pennsylvania doctoral graduate student studying linguistics and a software technologist interested in civic technology.

Joshua Tauberer, 26, currently splits his time between two fields that occasionally overlap:

What he intends to focus on full-time after the completion of his Ph.D. is the use of software technology to support civic life. He entered this field by creating the website www.GovTrack.us in 2004, which is a fully automated, nonpartisan and noncommerical website that tracks the activities of the U.S. Congress and compiles the only open database of U.S. legislative information. Joshua is currently a coordinator of The Open Senate Project, a project of the Sunlight Foundation, which will issue a report in 2009 making recommendations to U.S. Senate leadership about how the Congress can better use technology in the public's interest. Some of Joshua's recommendations in earlier The Open House Project report were incorporated into appropriations bill markup (2008). He has also helped to revise the OMB web policy best-practices guide on webcontent.gov.

He is currently, and primarily, a full-time "ABD" graduate student in the doctoral linguistics program at the University of Pennsylvania. The working title of his dissertation, expected to be completed in 2009, is "The acquisition of the complex realization of [voice] in American English," a study at the intersection of corpus phonetics, phonology, and language acquisition. The dissertation intends to contribute to the field's understanding of the interplay of articulatory gestures involved in the production of speech sounds, especially during their acquisition in children. Joshua has also studied natural language syntax and semantics, some machine learning, and computational modeling of language acquisition in the framework of Optimality Theory, the subject of his master's thesis.

Joshua is also interested in the use of the technology behind the Semantic Web to create interconnected databases of political information. Such a system would foster computer-aided education as well as government oversight by making it cheaper to find connections between entities in disparate parts of the political world.

Education

Ph.D. in progress, Linguistics, University of Pennsylvania **2009 (expected)**

A.M., Linguistics, University of Pennsylvania **2008**

Dissertation in progress on "The acquisition of the complex realization of [voice] in American English." Subfields of focus: corpus phonetics, formal syntax, semantics, computational modeling of language learning. Teaching assistant for Intro. to Cognitive Science, Intro. to Linguistics, graduate-level Intro. to Phonetics. Editor, U. Penn Working Papers in Linguistics.

A.B., Princeton University **2004**

Bachelor's degree (A.B.) in psychology, certificates (~minors) in linguistics, applications of computing. Also was news/features editor of The Daily Princetonian.

Experience

Creator — GovTrack.us **2003–Present**

Created the well-respected website http://www.govtrack.us which tracks the activities in the U.S. Congress, providing a free, citizen-oriented reference and tracking service covering the activities of Congress. Webby Award nominee (2006), Technorati Developer's Contest winner (2005), covered in The New York Times, mentioned in The Washington Post. 20,000 daily visitors.

Co-Coordinator — The Open Senate Project **2008–Present**

To issue a report to U.S. Senate leadership with the encouragement of majority leader Harry Reid regarding use of technology to improve legislative transparency. A project sponsored by the Sunlight Foundation. Also contributed to the Open House Project report in 2007 submitted to U.S. House of Representatives leadership, following the encouragement of Speaker Nancy Pelosi. See http://www.theopenhouseproject.com, http://www.theopensenateproject.com.

Chief Software Architect — LARSA, Inc. (www.larsa4d.com) **2004–Present**

Advanced software package for structural engineers, primarily used on bridge and earthquake analysis and design. Scientific programming, user interface programming, technical writing.

Freelance Writer — XML.com **2006**

Wrote three articles on the Semantic Web/RDF, one an introduction ("What is RDF"), and two making use of data produced by the U.S. government ("Public Data & The Semantic Web", "Query Census Data with RDF") under the heading "Hacking Congress". See http://xml.com/pub/at/33.

Invited Talks (Government Transparency)

Oct. 11, 2008. *Civic Hacking*. Presented in the Politics & Transparency panel of the Free Culture 2008 Conference at Berkeley. (Note: "hacking" here means "programming", not "breaking in".)

Jan. 15, 2008. *Open government data policy and a semantic future for civics*. Presented in Civics in the Cloud panel at "Computing in the Cloud", Center for InfoTech Policy, Princeton University.

Conference Publications (Linguistics)

2009. *Goldilocks Meets the Subset Problem: Evaluating Error Driven Constraint Demotion (RIP/CD) for OT Learning*. Presented at the 32nd Annual Penn Linguistics Colloquium, February 23, 2008. To appear in the U. Penn Working Papers of Linguistics 15.1, 2009, Proceedings of the 32nd Penn Linguistics Colloquium.

2008. *Predicting Intrasentential Pauses: Is Syntactic Structure Useful?* In Barbosa, P. A., Madureira, S., and Reis, C. (Eds.) Proceedings of the Speech Prosody 2008 Conference, May 6-9, 2008. Campinas, Brazil: Editora RG/CNPq. Pp 405-408.

2007. with L. Champollion. and M. Romero. *The Penn Lambda Calculator: Pedagogical Software for Natural Language Semantics*, in T. Holloway King and E. M. Bender (eds.), Proceedings of the Grammar Engineering across Frameworks (GEAF) 2007 Workshop. CSLI On-line Publications.

Other Technology Projects

SemWeb C# Library for RDF (2005–present); Thunderbird Extension for Sender Verification (2004–present); The Penn Lambda Calculator: A semantics pedagogical tool (2006–2007); Praat-Py: Phonetics tool extension (2007).

Skills

Programming: C#, Python, Java, Perl, XML, XSLT, RDF
Mathematics: Linear algebra, sparse systems, dimensionality reduction, machine learning