Nathaniel J. Smith

CONTACT

Phone: +1 510-332-7727
Email: `njs@pobox.com`
Web: `https://vorpus.org`

DEGREES

B.A. (with High Honors), Cognitive Science and Mathematics, University of California Berkeley, 2003. (*Valedictorian: Department of Mathematics*).
Honors thesis title: Gesture and beyond
Thesis supervisors: Eve Sweetser and Dan Slobin.

Ph.D. University of California San Diego, Cognitive Science, 2011.
Dissertation title: Scaling up psycholinguistics
Advisors: Roger Levy and Marta Kutas.

CAREER

Postdoctoral researcher, School of Informatics, University of Edinburgh, 2011-2015.
Mentor: Mark Steedman.

Computational fellow (assistant research scientist), UC Berkeley Institute for Data Science, 2015-2018.

Freelance software developer, 2018-present.

PUBLICATIONS UNDER REVIEW OR IN PRESS

Smith, N. J. (submitted). ZS: A file format for efficiently distributing, using, and archiving record-oriented data sets of any size.

PEER REVIEWED PUBLICATIONS

Abend, O., Kwiatkowski, T., **Smith, N. J.**, Goldwater, S., & Steedman, M. (2017). Bootstrapping Language Acquisition. *Cognition, 164*, 116-143.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology, 52*(2), 157-168.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology, 52*(2), 169-189.

Smith, N. J., Goodman, N. D., & Frank, M. C. (2013). Learning and using language via recursive pragmatic reasoning about other agents. In C. J. C. Burges and L. Bottou and M. Welling and Z. Ghahramani and K. Q. Weinberger (Eds.), *Advances in neural information processing systems (NIPS) 26* (pp. 3039-3047).

Burns, M. D., Bigdely-Shamlo, N., **Smith, N. J.**, Kreutz-Delgado, K., & Makeig, S. (2013). Comparison of Averaging and Regression Techniques for Estimating Event Related Potentials. In *IEEE Engineering in Biology and Medicine Conference*, Osaka, Japan.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302-319.

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1637-1642). Austin, TX: Cognitive Science Society.

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1313–1318). Austin, TX: Cognitive Science Society.

Smith, N. J., Chan, W., & Levy, R. (2010). Is perceptual acuity asymmetric in isolated word recognition? Evidence from an ideal-observer reverse-engineering approach. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1483–1488). Austin, TX: Cognitive Science Society.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: a formal model and empirical investigation. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 595–600). Austin, TX: Cognitive Science Society.

BOOK CHAPTERS

Smith, N. J. (in press). Blending across modalities in mathematical discourse. In L. D. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives in gesture and embodiment in mathematics*. Charlotte, NC: Information Age Publishing.

Kutas, M., DeLong, K., & **Smith, N. J.** (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain* (pp. 190–207). Oxford, Oxford University Press.

SCIENTIFIC SOFTWARE

Smith, N. J., van der Walt, S., Zhao, H., and others (2015-2017): *viscm*, a principled tool for analyzing and designing high-quality colormaps for data visualization.

Smith, N. J., Futrell, R., and others (2016): *colorspacious*, a library for performing colorspace conversions and estimating perceptual properties of color stimuli.

Smith, N. J. (2014): *ZS*, a fast and space-efficient database format for distributing, using, and archiving arbitrarily large record-oriented datasets, such as the Google n-grams.

Smith, N. J. (2013): *rERPy*, a Python toolkit for doing ERP/ERF and rERP/rERF analysis of brainwave data.

Smith, N. J. (2012): *Patsy*, a Python library for describing statistical models (especially linear models, or models that have a linear component) and building design matrices. Patsy brings the convenience of R "formulas" to Python.

I'm also currently (since 2012) a core developer for Numpy, the main numerical package for Python, and the author of Python Enhancement Proposal 465, which gave the Python language a dedicated notation for matrix multiplication (`mat1 @ mat2`).

Since 2018, I've also been a core developer for Python itself.

PEER-REVIEWED PRESENTATIONS WITHOUT PROCEEDINGS

Smith, N. J.. (2016, July). *Reinventing the .whl: New Developments in the Upstream Python Packaging Ecosystem* Talk presented at SciPy 2016, Austin, TX.

Smith, N. J. and van der Walt, S. (2015, July). *How we designed a new default colormap for Matplotlib (and you can too).* Talk presented at SciPy 2015, Austin, TX.

Smith, N. J., & Levy, R. (2011, March). *Comparing cloze versus corpus probabilities in self-paced reading.* Poster presented at the CUNY conference on sentence processing, Palo Alto, CA.

Smith, N. J., & Levy, R. (2010, March). *Bias in the cloze task.* Poster presented at the CUNY conference on sentence processing, New York, NY.

Smith, N. J., & Levy, R. (2010, March). *Fixation durations in first-pass reading reflect uncertainty about word identity.* Poster presented at the CUNY conference on sentence processing, New York, NY.

Smith, N. J., & Narayan, S. (2008, October). *Fidgeting is not random: Rhythmic leg motion, speech, and gesture.* Oral presentation at Conceptual Structure, Discourse, and Language (CSDL), Cleveland, OH.

Smith, N. J., & Levy, R. (2008, October). *Probabilistic prediction and the continuity of language processing.* Oral presentation at Conceptual Structure, Discourse, and Language (CSDL), Cleveland, OH.

Smith, N. J. (2008, March). *Surprisal as optimal behavior: a formal model and empirical investigation.* Oral presentation at the CUNY conference on sentence processing, Chapel Hill, NC.

Smith, N. J. (2007, July). *Persistent structure: A unifying concept in discourse.* Oral presentation given at the International Cognitive Linguistics Conference (ICLC), Kraków.

Smith, N. J., & Narayan, S. (2007, June). *Fidgeting is not random: Rhythmic leg motion, speech, and gesture.* Poster presented at Integrating Gestures, Evanston, IL.

Smith, N. J. (2007, June). *Gesture without interaction: Cognitive uses for a communicative capacity. Oral presentation at Integrating Gestures, Evanston, IL.*

Smith, N. J., and Núñez, R. (2006, November). *Speaking for thinking*. Oral presentation at Conceptual Structure, Discourse, and Language (CSDL), San Diego, CA.

Narayan, S., **Smith, N. J.**, and Sweetser, E. (2005, January). *Constructional histories:* coming across evidence for grammatical theory. Oral presentation at the annual meeting of the Linguistic Society of America, Oakland, CA.

Smith, N. J. (2004, October). *'Buoys' in co-speech gesture and ASL.* Oral presentation at Conceptual Structure, Discourse, and Language (CSDL), Edmonton.

WORKSHOPS ORGANIZED

Organized and ran the Python Compilers Workshop, 11-12 July 2016, UT Austin (co-located with SciPy 2016).

A bayesian model of language learning, use, and emergence through interaction. Institute for Language, Cognition and Computation/Human Communication Research Center (ILCC/HCRC) seminar series, School of Informatics, University of Edinburgh, 23 January 2015.

Predictability and probability in language comprehension. Baayen lab meeting, Eberhard Karls Universität Tübingen, 27 March 2014.

rERP: Regression-based estimation of ERP waveforms. Baayen lab meeting, Eberhard Karls Universität Tübingen, 27 March 2014.

Building a Bayesian bridge between the physics and the phenomenology of social interaction.

- Tenenbaum lab meeting, BCS, MIT, 12 February 2014.
- Center for Language Sciences colloquium, BCS, University of Rochester, 14 February 2014.
- Ling Lang Lunch, CLPS, Brown University, 19 February 2014.
- TedLab, BCS, MIT, 25 February 2014.
- Department of Psychology, UMass Amherst, 26 February 2014.

Predictability and probability in language comprehension. *Mayfest 2013 - Linguistically Predictable: When, How, and Why Do We Predict in Language?*, University of Maryland, 4 May 2013.

rERP: Estimating ERPs using statistical regression. Presented at:

- TedLab, MIT, 15 May 2013.
- University of Maryland, 2 May 2013.
- Knight lab meeting, UC Berkeley, 19 February 2013.

Probabilistic prediction in human language processing. Gahl lab meeting, UC Berkeley, 11 February 2013.

Reconciling speech and gesture with action and perception using a bayesian phenomenology. Berkeley Gesture Group, UC Berkeley, 1 February 2013

Probabilistic prediction in human language processing. Institute for Language, Cognition and Computation/Human Communication Research Center (ILCC/HCRC) seminar series, School of Informatics, University of Edinburgh, 5 October 2012.

HONORS AND AWARDS

Center for Research in Language (CRL) Pre-Doctoral Fellowship (Funded by NIH training grant T32-DC000041), 2009–2010

Institute for Neural Computation (INC) Pre-Doctoral Fellowship (Funded by NIH training grant T32-MH20002), 2008–2009

NSF Graduate Fellowship, 2004–2009

Students supervised

Qian, Benben (2013). *Comparing human and computer predictions of upcoming words in text*. MSc in Informatics, University of Edinburgh. (Sole supervisor)

Tutorials

With Roger Levy and Klinton Bicknell, presented: *Computational psycholinguistics: Integrating NLP modeling and experimental psycholinguistics to investigate real-time human language use*. Tutorial at the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT) 2010 conference, Los Angeles, CA, 1 June 2010.

University courses

Teaching Assistant, Learning, Memory and Attention, UC San Diego (Winter 2004, Spring 2009)

Teaching Assistant, Distributed Cognition, UC San Diego (Fall 2005, Fall 2007)

Teaching Assistant, Modeling and Data Analysis, UC San Diego (Fall 2006)

Teaching Assistant, Design and Analysis of Experiments, UC San Diego (Winter 2008)

Other

Mentor/Administrator of Google Summer of Code students for the Monotone project (Summer 2005, Summer 2006)

SERVICE

Ad hoc journal article reviewer for *Journal of Memory and Language*, *Psychophysiology*, *Language and Speech*, *Transactions of the Association for Computational Linguistics*.

Conference reviewer for the annual meeting of the Association for Computational Linguistics (ACL), the annual meeting of the Cognitive Science Society, the annual meeting of the European Chapter of the Association for Computational Linguistics (EACL), the annual meeting of the North American chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), the annual Cognitive Modeling and Computational Linguistics (CMCL) workshop.