Cyrus Shaoul

1109 Black Mountain Rd.
Dummerston, VT
USA, 05301

Phone: 802-254-4520
Email: cyrus@leela.ai

Education

University of Alberta, Edmonton, AB. Ph.d. in Psychology, completed in June 2012. Doctoral Dissertation Title: *The Processing of Lexical Sequences*
Downloadable at: http://hdl.handle.net/10402/era.26026

University of Alberta, Edmonton, AB. M.Sc. in Psychology, completed in June 2008. Master's Thesis Title: *Optimizing HAL parameter space for predicting lexical access and semantic decision latency.*

Massachusetts Institute of Technology, Cambridge, MA. B.S. in Brain and Cognitive Science, graduated in 1993.

Employment

CEO — Leela AI Inc.
Newton, MA — 2017–current
Co-founded a small startup taking a new, hybrid AI approach to the area of intelligent agents.

Technology Consultant — Tomigaia Inc.
New York, NY and Edmonton, AB — 2001–2004
Owned and operated a one-person consulting company specializing in IT technology transfer. Did extensive contract work for NTT on VOIP technology related to security and standardization. Attended IETF meetings and participated in VOIP related IETF working groups.

CTO — Digital Garage Inc.
Tokyo, Japan — 1994–2001
Co-founded Japan's first web technology company with Joichi Ito (the current director the of MIT Media Lab) and two other co-founders. Managed technology development for Internet-scale search and other web applications. Helped localize and launch Infoseek Japan, Japan's first search engine. Worked with many Japanese and international clients (IBM Japan, CSK, TIS, Charles Schwab, and many others). Was personally responsible for corporate technology strategy development, negotiations with customers, hiring, investor relations and many other activities. Grew company from 5 employees to 120 employees during this period. Member of the Board of Directors until the completion of IPO on the Tokyo stock exchange. Left Japan after IPO to raise family in the USA. Digital Garage continues to be a leader in web-related technologies in Japan and other parts of Asia.

Researcher — NTT Communications Laboratory
Yokusuka, Japan — 1993–1994
Software developer at the NTT human interface research laboratory. Worked on a large-scale networked immersive 3-D communication system, a similar to the commercial products Second Life and World-of-Warcraft, which it predated.

Skills

- Languages: English (native), fluent French and Japanese, semi-fluent Spanish and German.
- Computer Languages: Python, Lisp/Scheme, Java, C++, C, Javascript, R, Matlab and others
- Operating Systems: Unix (Linux, local or cloud-based), MacOS X and others. Beowulf clusters and other batch-processing supercomputer systems.
- Scientific Software: R (including package development), knitr, CLUTO, Gnuplot, SPSS, SYSTAT and others.
- Machine Learning (R packages): boosted regression trees (gbm), random forests (party), deep learning (MXNetR, darch, deepnet, H2O and deepr)
- Statistical Methods (R package): Multiple regression, (generalized) linear mixed-effects regression models (lme4), generalized additive mixed models (mgcv), ordinal regression (ordinal), visualization (ggplot2, lattice), clustering (GMM), randomization tests, bootstrapping, inference by model comparison and others.
- Eye-tracking Systems: SR Research Eyelink2 and Eyelink1000(plus)
- EEG Systems: BioSemi ActiveTwo, Cognionics Quick20
- Optical Brain Imaging Systems: BioPac FNIR100C

Academic Employment

Landmark College Institute for Research and Training — Putney, VT
Associate Professor, Senior Academic Researcher — July 2015–June 2017
Did research on the cognitive neuropsychology of learning using eye-tracking, EEG and fNIRS neuroimaging techniques. Designed research studies and analyzed datasets using the R statistical environment.

University of Tübingen, Dept. of Linguistics — Tübingen, Germany
Postdoctoral Fellow — Sept 2012–June 2015
Worked in Dr. Harald Baayen's Quantitative Linguistics group. Did research on new models of lexical processing based on the theory of naive discriminative learning (NDL).

University of Alberta, Dept. of Psychology — Edmonton, AB, Canada
Research Assistant — Sept 2004–June 2012
Worked with Professor Chris Westbury on investigating the HAL model of lexical semantic memory. Implemented a program called HiDEx (High Dimensional Explorer) to test hypotheses about HAL parameters and semantic processing. During doctoral work, investigated n-gram effects in reading. Designed and implemented psycholinguistic experiments. Analyzed and performed statistical analyses on data collected.

Teaching Experience

- Landmark College
  - Co-instructor (with Dr. R. Bryck), Self-Regulation, Motivation and Student Engagement (Graduate Level course, taught completely online)
- University of Tübingen

  - Instructor, Introduction to cognitive models of language processing (Co-taught with Prof. R. H. Baayen)
  - Guest Lecturer, Introduction to Linguistics for Cognitive Science Majors (Taught by Prof. R. H. Baayen)
- University of Alberta
  - TA, Advanced Topics in Development: Children and the Web (4th year undergraduate class), 2009
  - TA, Advanced Methods: Monte Carlo Methods in Statistics (4th year undergraduate class), 2008, 2010, 2011
  - TA, Learning & Behaviour, University of Alberta, 2007
  - Lab TA, Design & Analysis of Experiments in Psychology (4th year undergraduate class), University of Alberta, 2005
  - TA, Memory and Cognition (4th year undergraduate class), University of Alberta, 2005, 2007

Book Chapters

Shaoul, C. & Westbury, C. (2011) HiDEx: The High Dimensional Explorer in *Applied Natural Language Processing and Content Analysis: Identification, Investigation, and Resolution* (McCarthy, P. and Boonthum, C. Eds.) IGI Global, Hershey, PA, USA
http://www.igi-global.com/book/applied-natural-language-processing/49572

Published Papers

Tse, C. S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior research methods*, 49(4), 1503-1519.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, 31(1), 106-128.

Westbury, C., Shaoul, C., Moroschan, G., & Ramscar, M. (2016). Telling the worldÕs least funny jokes: On the quantification of humor as entropy. *Journal of Memory and Language*, 86, 141-156.

Shaoul, C., Baayen, H. R., Westbury, C. F. (2015) N-gram probability effects in a cloze task. *Journal of the Mental Lexicon, Bridging the Methodological Divide: Linguistic and psycholinguistic approaches to formulaic language*, pp. 437-472, 9:3
http://www.ingentaconnect.com/content/jbp/ml/2014/00000009/00000003/art00005

Ramscar, M. , Hendrix, P. , Shaoul, C. , Milin, P. & Baayen R. H. (2014) The Myth of Cognitive Decline: Non-linear Dynamics of Lifelong Learning *Topics in Cognitive Science*, Early view.
http://onlinelibrary.wiley.com/doi/10.1111/tops.12078/abstract

Shaoul, C. , Westbury, C. & Baayen, R. H. (2013) The subjective frequency of word n-grams. *Psihologija*, 46, 497-537.
http://www.doiserbia.nb.rs/journal.aspx?issn=0048-5705

Westbury, C. , Shaoul, C. , Hollis, G. , Smithson, L. , Briesemeister, B. B. , Hofmann, M. J. & Jacobs, A. M. (2013) Now you see it, now you don't: On emotion, context, & the algorithmic prediction of human imageability judgments *Frontiers in Language Sciences*
http://www.frontiersin.org/Journal/10.3389/fpsyg.2013.00991/abstract

Gries, S. Th., Newman, J. & Shaoul, C. (2011) N-grams and the clustering of registers. *Empirical Language Research.* 4:1
http://ejournals.org.uk/ELR/article/2011/1

Shaoul, C. & Westbury, C. (2011) Formulaic sequences: Do they exist and do they matter? Methodological and Analytic Frontiers in Lexical Research (Part II). Special Issue of The Mental Lexicon 6:1, Westbury, Chris (ed.)
http://dx.doi.org/10.1075/ml.6.1.07sha

Shaoul, C. & Westbury, C. (2010) Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods* , 42:2, 393-413
http://dx.doi.org/10.3758/BRM.42.2.393

Westbury, C., Hollis, G., & Shaoul, C. (2007) LINGUA: Language Independent of Neighborhood Generator of the University of Alberta. *Journal of the Mental Lexicon*, 2:2, 273-286
http://tinyurl.com/92z6wgz

Shaoul, C. & Westbury, C. (2006) Word frequency effects in high-dimensional co-occurrence models: A new approach. *Behavior Research Methods*. 38:2, 190-195
http://dx.doi.org/10.3758/BF03192768

Refereed Contributions

2014

Shaoul, C., Willits, J. (2014) Modeling multiple aspects of language acquisition using Naive Discrimination Learning *In Proceedings of Society for Computers in Psychology SCiP 2014* (Long Beach, CA, November, 2014). [talk]

Shaoul, C., Milin, P., Hendrix, P., Arppi, A., Ramscar, M. & Baayen, R. H. (2014) NDL: An R package for large-scale naïve discriminative learning. 2014 Psychometric Computing Conference, Tübingen. [talk]

Shaoul, C., van Rij, J., Baayen, R. H., Milin, P. (2014) Word Comprehension in Reading is Fast and Furious. 9th International Conference on the Mental Lexicon, Niagara-on-the-lake, Canada. [poster]

2013

Shaoul, C., Milin, P., Hendrix, P., Arppi, A., Ramscar, M. & Baayen, R. H. (2013) NDL: An R package for large-scale naïve discriminative learning. 43rd annual meeting of the Society for Computers in Psychology, Toronto, Canada. [talk]

Shaoul, C., Milin, P., Hendrix, P., Ramscar, M. & Baayen, R. H. (2013) The consequences of accumulating experience for lexical processing. 8th International Morphological Processing Conference,

Cambridge, UK. [talk]

Shaoul, C., Milin, P., Hendrix, P., Arppi, A., Ramscar, M. & Baayen, R. H. (2013) The NDL package: a tool for building highly scalable learning models. What Can We Do With 500 Billion Words? Conference, Bloomington, IN, USA [talk]

Shaoul, C. , Baayen, R. H. , Westbury, C. (2013) Agreeing with Google: We are sensitive to the relative corpus frequency of *N*-grams.12th International Cognitive Linguistics Conference, Edmonton, Canada [talk]

2010

Westbury, C. & Shaoul, C. (2010). Co-occurrence measures predict human concreteness judgments. 51st Annual Meeting of The Psychonomic Society, St. Louis, MO, USA. [poster]

Shaoul, C. , Westbury, C. & Baayen, H. (2010) Whole Phrase Frequency Predicts Behavior In a Cloze Task. Presentation, 7th International Conference on the Mental Lexicon, Windsor, Ontario, July 3rd, 2010 [talk]

2009

Shaoul, C. , Westbury, C. & Baayen, H. (2009) Agreeing with Google: We are Sensitive to the Relative Orthographic Frequency of Phrases. Poster, 50th Annual meeting of the Psychonomic Society, November 22, 2009, Boston, MA USA [poster]

Shaoul, C. & Westbury, C. (2009) Measuring contextual similarity using HiDEx. *In Proceedings of Society for Computers in Psychology SCiP 2009* (Boston, November, 2009). [poster]

Shaoul, C. , Westbury, C. & Baayen, H. (2009) *Agreeing with Google: We are Sensitive to the Relative Corpus Frequency of Phrases.* Presentation, 10th Annual meeting of the American Association for Corpus Linguistics , Edmonton, Alberta, October 9th, 2009 [talk]

Gries, S. Th., Newman, J. , Shaoul, C. & Dilts, P. (2009) N-grams and the clustering of genres. Corpus, collocation, register variation workshop at the 31st Annual Meeting of the *Deutsche Gesellschaft für Sprachwissenschaft*, March 6th 2009 [talk]

2008

Shaoul, C. & Westbury, C. (2008) Performance of HAL-like word space models on semantic categorization tasks. Proceedings of the Workshop on Lexical Semantics, ESSLLI 2008. 42-46 [talk]
http://wordspace.collocations.de/doku.php/workshop:esslli:proceedings

2007

Shaoul, C. & Westbury, C. (2007) Walking in space: Optimizing parameter settings in co-occurrence models of meaning. Society for Computers in Psychology Symposium on Co-occurrence and Lexical Organization, Long Beach, CA Nov 15th, 2007. [talk]

2005

Shaoul, C. & Westbury, C. (2005) Towards a more psychologically relevant high dimensional model

of lexical semantics. *In Proceedings of Society for Computers in Psychology SCiP 2005* (Toronto, November, 2005). [talk]

Invited Contributions

2015

Shaoul, C. and Ramscar, M. (2015) Discriminative learning models. *Building Realistic Models of Language Development Workshop* Manchester, UK. [talk]
http://www.psych-sci.manchester.ac.uk/ldd/lucid/events/modellingworkshop/

2014

Shaoul, C., & Ramscar M., (March 2014) Building Rescorla-Wagner models from mega-corpora using naive discriminative learning. Workshop on using corpora for quantitative and psycholinguistic analysis. Eszterhàzy College, Eger, Hungary.
http://iqma.xyz/workshops/eger.html

Non-refereed Contributions

2015

Ramscar, M., Shaoul, C., & Baayen, R. H. (2015). Why many priming results don't (and won't) replicate: A quantitative analysis. Manuscript, University of Tübingen.

2014

Westbury, C. , Shaoul, C. , Hollis, G. , Smithson, L. , Briesemeister, B. B. , Hofmann, M. J. & Jacobs, A. M. (2014) Human judgments estimated from co-occurrence with affect terms.
http://www.psych.ualberta.ca/˜westburylab/downloads/AffectEstimates.download.html

2013

Shaoul, C., Arppi, A. , Milin, P., Hendrix, P. & Baayen, R. H. (2013) *The NDL package for R* Source code and documentation on the Comprehensive R Archive Network (CRAN)
http://cran.r-project.org/web/packages/ndl

Shaoul, C., Bitschnau, S., Schilling, N. , Milin, P., Hendrix, P. & Baayen, R. H. (2013) *The NDL2 package for R* (under development)

Shaoul, C. & Westbury, C. (2013) *HiDEx Beta version 0.092* Open-source tool implementing the HAL model.
http://www.psych.ualberta.ca/˜westburylab/downloads/HiDEx.download.html

2011

Shaoul, C. (2011) *Phrase frequency: a new perspective on language processing.* Departmental Seminar, Department of Psychology, University of Alberta, March 25th, 2011 [talk]

2010

Shaoul, C. & Westbury, C. (2010) *Neighborhood Density Measures for 57,153 English Words.* http://www.psych.ualberta.ca/˜westburylab/downloads/westburylab.arcs.ncounts.html

Shaoul, C. & Westbury, C. (2010) *Whole phrase frequency predicts behavior in a cloze task.* Presentation, 24th Royce Conference, University of Alberta, March 19th, 2010 [talk]

Shaoul, C. Westbury, C. & Baayen, H. (2010) *Agreeing with Google: We are Sensitive to the Relative Corpus Frequency of Phrases.*, 28th Annual Banff Annual Seminar In Cognitive Science, Banff, Alberta, May 2nd, 2009 [poster]

Shaoul, C. & Westbury, C. (2010) *A anonymized multi-billion word USENET corpus (2005-2010)* http://www.psych.ualberta.ca/˜westburylab/downloads/usenet.download.html

2009

Shaoul, C. Westbury, C. & Baayen, H. (2009) *Agreeing with Google: We are Sensitive to the Relative Corpus Frequency of Phrases.* 23rd Royce Conference, University of Alberta, March 20th, 2009 [talk]

2007

Westbury, C. & Shaoul, C. (2007) *Orthographic Neighborhoods for over 111,000 English words.* Edmonton, AB: University of Alberta.
http://www.psych.ualberta.ca/˜westburylab/downloads/ON.download.html

Shaoul, C. & Westbury, C. (2007) Optimizing HAL parameter space for predicting lexical access. Royce Conference, March 2007 & the Symposium on Language and Memory at Manitoba, Winnipeg, April 2007. [poster]

2006

Shaoul, C. & Westbury, C. (2006) An investigation into the effect of changing the parameters of the HAL model. Mental Lexicon Conference, Montreal, Oct, 2006. [poster]

Shaoul, C. & Westbury, C. (2006) Global Co-occurrence Density Predicts Both Lexical and Semantic Decision RTs. CSBBCS conference Saskatoon, June , 2006. [talk]

Shaoul, C. & Westbury, C. (2006) An investigation into the effect of changing the parameters of the HAL model. 2006 Mental Lexicon Conference. [poster]

Westbury, C. Hollis, G. & Shaoul, C. (2006) *LINGUA: The Language Independent of Neighborhood Generator of the University of Alberta*.
http://www.psych.ualberta.ca/ westburylab/downloads/lingua.download.html

Shaoul, C. & Westbury, C. (2006) *USENET Orthographic Frequencies for the 40,481 words in the English Lexicon Project.*
http://www.psych.ualberta.ca/ westburylab/elp.download.html

Shaoul, C. & Westbury, C. (2006) *USENET Orthographic Frequencies for 1,618,598 types.*.

http://www.psych.ualberta.ca/ westburylab/wlallfreq.download.html

Shaoul, C. & Westbury, C. (2006) *Global Co-occurrence Density Predicts Both Lexical and Semantic Decision RTs*. 2006 CSBBCS conference. [poster]

Shaoul, C. & Westbury, C. (2006) *Global Co-occurrence Density Predicts Both Lexical and Semantic Decision RTs*. 20th Royce Conference, University of Alberta, 17th of March, 2006. [poster]

2005

Shaoul, C. & Westbury, C. (2005) Influence of Orthographic Frequency of Words on the HAL Model of Semantic Space. 2005 Royce Conference [poster]

Awards

- Canadian Psychology Association Doctoral Dissertation Award, 2012
- Tolman Undergraduate Teaching Award, 2011
- J. Gordin Kaplan Graduate Student Award (Travel Grant), 2007
- Walter H. Johns Graduate Fellowship (3 year, $4,949 per year), 2007-2010
- NSERC (Natural Sciences and Engineering Research Council of Canada) PGS-D (3 year doctoral scholarship, $21,000 per year), 2007-2010
- Castellan Award, November, 2005 (Awarded to the outstanding student paper presented at the annual SCiP conference)
- Alberta Advanced Education Achievement Scholarship, awarded in December 2005 by the Alberta Heritage Scholarship Fund.

Service

- Member of the 2013 Program Committee for EMNLP 2013 : Conference on Empirical Methods in Natural Language Processing
- Member of the 2012 Program Committee for SEM: The First Joint Conference on Lexical and Computational Semantics
- Served as a reviewer for the Cognitive Science Society's annual conference on multiple occasions.
- Past Psychology department representative to the GSA (Graduate Student Association) Council
- Past Graduate student member of the FGSR (Faculty of Graduate Studies and Research) academic appeals committee.
- Past Graduate Student Association representative on the FGSR Council
- Past Chairman of the Psychoquium Committee at the University of Alberta http://www.psych.ualberta.ca/˜psychoqu/
- Served as a reviewer for the journals *Cognition*, *The Mental Lexicon*, *Cognitive Science*, *Cognitive Processing* and *Behavior Research Methods*.

- Community Service: Worked with the Outreach office of the Faculty of Graduate Studies and Research on Community Outreach projects: ⋆ Supporting the Belgravia School Science Fair (1^{st}-6^{th} graders) (2004) ⋆ Read-In at the Oliver School (Kindergarden and 9^{th} graders) (2005) ⋆ Guest Lecturer at the Ottwell School (7^{th} graders) (2006) ⋆ Debate judge at the Kenilworth School (9^{th} graders) (2006) ⋆ Read-In at Allendale and Belgravia Schools (2nd, 7^{th}, 8^{th} and 9^{th} graders) (2006) ⋆ Read-In at Grandin Schools (4^{th} and 5^{th} graders) (2007) ⋆ Mentoring a student from WP Wagner High School through the F.I.R.S.T. (Focus in Research, Science and Technology) Honours Program (2007) ⋆ Judge at the Alberta Lego Robot Competition (2008)
- Webmaster of the Westbury Lab Web Site (2004-2012)