General Examinations Proposal

Soroush Vosoughi

Feb 29, 2012

Introduction

This exam is motivated by one general question, how can we build machines that learn to communicate as humans do? Specifically, how can we build intelligent agents that can perceive their world and communicate meaningfully with other agents (artificial or otherwise)?

In order to answer this question, we start by trying to understand and model how children learn to use language by studying theoretical and empirical work in the field of child language acquisition. We then look at philosophical, cognitive and computational theories on how meaning is derived from sensory-motor signals (which are the agents only connections to the world) and how that meaning is abstracted and represented in our mental models. We finish by looking at previous attempts at building artificial agents that implement these concepts and study their successes and failures.

Main Area: Grounded Semantics

Examiner
Prof. Deb Roy
Associate Professor of Media Arts and Sciences
Director of the Cognitive Machines Group
MIT Media Laboratory

Description
The reading list for the main area examines theories of multi-modal grounded language acquisition from the perspectives of philosophy, cognitive science and artificial intelligence. The reading list covers a wide variety of literature in grounded semantics and related fields, from established and foundational works to new work in the area. Specifically, this list covers theories and models of child language acquisition, knowledge representation, embodied agents and situated semantics.

Requirement
The completion requirement for this area will be a publishable-quality paper, evaluated by Prof. Deb Roy.

Examiner's signature: Date:

Reading List
Foundations

Ballard, D.H., Animate vision, Artificial Intelligence Journal 48, 57-86, 1991

Dennett, D. Evolution, Error and Intentionality. In The Intentional Stance, pages 287–321. The MIT Press, 1987.

Dennett, D. Brainchildren: Essays on Designing Minds, chapter 5: "Real Patterns", pages 95–120. The MIT Press, 1998.

Drescher, G. Made-Up Minds: A Constructivist Approach to Artificial Intelligence. The MIT Press, 1991.

Dreyfus, H. From Micro-Worlds to Knowledge Representation: AI at an Impasse In Mind Design II (ed. J. Haugeland), 1997

Glenberg, and Arthur C. Graesser, editors, Symbols and Embodiment: Debates on Meaning and Cognition, chapter 11. Oxford University Press.

Fauconnier, G.: 1997, Mappings in Thought and Language, Cambridge University Press, Cambridge.

Harnad, S. The symbol grounding problem. Physica D, 42:335–346, Jun 1990.

Lock, J. An Essay Concerning Human Understanding (1690).

Minsky, M.: 1974, A Framework for Representing Knowledge. MIT AI Laboratory Memo 306, Massachusetts Institute of Technology, Cambridge, MA, USA.

Quine, W.V Word and Object. Cambridge, MA: MIT Press. 1969

Roy, D.: 2008, A mechanistic model of three facets of meaning. In Manuel de Vega, Arthur M.

Searle, JR (1969). Speech acts: An essay in the philosophy of language. Cambridge University Press.

Winograd, T and Flores, F.: 1987 Understanding Computers and Cognition : A New Foundation for Design. Addison-Wesley Professional.

Wittgenstein, L. Philosophical Investigations. 1945.

Cognitive Science

Bartlett, E. (1978). The acquisition of the meaning of color terms: A study of lexical development. Recent advances in the psychology of language: Language development and mother-child interaction, 89-108.

Bates, E., with L. Benigni, I. Bretherton, L. Camaioni, & V. Volterra. (1979). The emergence of symbols: Cognition and communication in infancy. New York: Academic Press.

Bickhard, M . Language as an interaction system. New Ideas in Psychology, 25 (2), pp. 70-86. (2007)

Bloom, P. How Children Learn the Meanings of Words. 2000.

Bruner, J. (1985). Child's Talk: Learning to Use Language

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. Nature, 385, 813-815

Pinker, S. (1979). Formal models of language learning. Cognition, 7(3), 217-283

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. Cognitive Science: A Multidisciplinary Journal, 29(6), 819-865.

Tomasello, M. (1995). Beyond Names for Things: Young Children’s Acquisition of Verbs. Lawrence Elbaum.

Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press.

Waxman, S., & Gelman, S. (2009). Early word-learning entails reference, not merely associations. Trends in cognitive sciences.

Yang, C. (2004). Universal Grammar, statistics or both? Trends in Cognitive Sciences, 8(10), 451-456.

Yu, C., Ballard, D., & Aslin, R. (2005). The role of embodied intention in early lexical acquisition. Cognitive Science: A Multidisciplinary Journal, 29(6), 961-1005.

Computational Models

Fazly, A., Alishahi, A., & Stevenson, S. (in press). A probabilistic computational model of cross-situational word learning. Cognitive Science.

Fleischman, M and Roy, D.: 2008, Grounded Language Modeling for Automatic Speech Recognition of Sports Video. HLT/NAACL. Columbus, OH.

Fleischman, M and Roy, D.: 2005, Why are verbs harder to learner than nouns? Initial insights from a computational model of situated word learning. 27th Annual Meeting of the Cognitive Science Society.

Fleischman, M and Roy, D.: 2005, Intentional Context in Situated Language Learning. Ninth Conference on Computational Natural Language Learning.

Gorniak, P and Roy, D.: 2004, Grounded Semantic Composition for Visual Scenes, Journal of Artificial Intelligence Research, Volume 21, pages 429-470

Gorniak, P and Roy, D.: 2005, Probabilistic Grounding of Situated Speech using Plan Recognition and Reference Resolution. Seventh International Conference on Multimodal Interfaces (ICMI 2005).

Gorniak ,P.: 2005 ,The Affordance-Based Concept, Peter Gorniak, Ph.D. Thesis, Massachusetts Institute of Technology.

Gorniak, P and Roy, D.: 2006, Perceived Affordances as a Substrate for Linguistic Concepts. Twenty-eighth Annual Meeting of the Cognitive Science Society, 6 pages.

Gorniak, P and Roy, D.: 2007, Situated Language Understanding as Filtering Perceived Affordances. Cognitive Science, 31(2), 197-231.

Jones, B., Johnson, M., & Frank, M. C. (2010). Learning words and their meanings from unsegmented child-directed speech. In Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Roy, D and Mukherjee., N.: 2005, Towards situated speech understanding: visual context priming of language models. Computer Speech & Language, 19(2):227–248.

Roy, D.: 2005, Semiotic Schemas: A Framework for Grounding Language in Action and Perception. Artificial Intelligence.

Roy, D.: 2005. Grounding words in perception and action: computational insights. Trends in Cognitive Science, 9(8), 389-396.

Roy, D.: 2003, Grounded Spoken Language Acquisition: Experiments in Word Learning. IEEE Transactions on Multimedia, 5(2): 197-209.

Roy, D.: 2002, A Trainable Visually-Grounded Spoken Language Generation System. In Proceedings of the International Conference of Spoken Language Processing.

Roy, D.: 2002, Learning Words and Syntax for a Visual Description Task.Computer Speech and Language.

Roy, D and Pentland, A.:2002, Learning words from sights and sounds: a computational model. Cognitive Science, 26(1):113–146.

Siskind, J. Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition. PhD thesis, Massachusetts Institute of Technology, 1992.

Siskind, J. M.: 2001, Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic, Journal of Artificial Intelligence Research 15, 31–90.

Xu, F and Tenenbaum, J. Word learning as bayesian inference. Psychological Review, 114(2):245–272, April 2007.

Yu, C and Ballard, D.H. A unified model of early word learning: Integrating statistical and social cues. Neurocomputing, 70(13-15):2149–2165, 2007.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross- situational statistics. Psychological Science, 18(5), 414-420.

Winograd, T. Understanding natural language, Cognitive Psychology, Volume 3, Issue 1, January 1972, Pages 1-191,

Contextual Area: Social Behavior and Intelligence

Examiner

Prof. Alex (Sandy) Pentland
Professor of Media Arts and Sciences
Director of the Human Dynamics Laboratory
MIT Media Laboratory

Description

The contextual area explores different forms of human communication and interaction, from speech to non-linguistic communication (also known as social signals) and influence models. The central theme of this area is to look at a communicative agent (be it human or artificial) not just as a self-contained input/output black box, but as an actor immersed in a social network, being influenced by(and in influencing) its peers through social signals.

Requirement

The written requirement for this area will be a 24 hour take home examination to be administered and evaluated by Prof. Alex Pentland.

Examiner's signature: Date:

Reading List

Computational Social Science:

Ballard, D.:1997, Deictic Codes for the Embodiment of Cognition, Behavioral and Brain Science.

Castellano, C and Loreto, V. 2009. Statistical physics of social dynamics. Reviews of Modern Physics, 81(2):591.

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. Cognitive Science: A Multidisciplinary Journal, 19(2), 233-263.

Dong, Wen. Modeling the Structure of Collective Intelligence. PhD thesis, Massachusetts Institute of Technology, 2010.

Lazer D, Pentland A, Adamic A, Aral S, Barabási A-L, et al. (2009) Life in the network: the coming age of computational social science. Science 323: 721

Pentland, A. 2006. Are We One? On the Nature of Human Intelligence. : 5 th Int'l Conf. On Development and Learning, Bloomington Il,

Language, Interaction and Social Signals

Basu, B., (2002) Conversational Scene Analysis. Unpublished doctoral thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA

Cassell, J. (2000). "More than Just Another Pretty Face: Embodied Conversational Interface Agents." Communications of the ACM 43(4): 70-78

Choudhury, T and Basu, S. 2004. Modeling conversational dynamics as a mixed memory markov process. In Proc. of Intl. Conference on Neural Information and Processing Systems (NIPS). Citeseer.

Delaherche E, Chetouani M. 2011. Automatic recognition of coordination level in an imitation task. In Proc. ACM J-HGBU Workshop

Dong W, Lepri B, Cappelletti A, Pentland AS, Pianesi F, and Zancanaro M. Using the influence model to recognize functional roles in meetings. In Proceedings of the 9th international conference on Multimodal interfaces, pages 271–278. ACM, 2007.

Dunbar, R., (1998) Grooming, Gossip, and the Evolution of Language. Cambridge, MA: Harvard University Pres

Georgiou P, Black M, Narayanan S. 2011. Behavioral signal processing for understanding (distressed) dyadic interactions: Some recent developments. In Proc. ACM J-HGBU Workshop.

Gorin, A. L., Riccardi, G. and Wright, J.: 1998, How may i help you?, Speech Communication.

Graf, H.P., Cosatto, E., Strom, V., Huang, F.J. (2002) Visual Prosody: Facial Movements Accompanying Speech, Fifth IEEE Int'l Conf. On Automatic Face and Gesture Recognition, Washington D.C., May 20-21, 2002.

Grice, Paul (1975). "Logic and conversation". In Syntax and Semantics, 3: Speech Acts, ed. P. Cole & J. Morgan. New York: Academic Press. Reprinted in Studies in the Way of Words, ed. H. P. Grice, pp. 22–40. Cambridge, MA: Harvard University Press (1989)

Grosz, B and Sidner C. Attention, intentions, and the structure of discourse. Computational Linguistics, 12(3):175–204, 1986.

Morenc, L. Computational Study of Human Communication Dynamics. In Proc. ACM J-HGBU 2011 Workshop

Nass, C., & Brave, S. (2004). Voice activated: How people are wired for speech and how computers will speak with us. Cambridge, MA: MIT Pres

Pan W, Dong W, Cebrian M, Kim T, Fowler JH, Pentland AS. 2011. Modeling Dynamical Influence in Human Interaction.

Pentland, A.; , "A Computational Model of Social Signaling" Pattern Recognition, 2006. ICPR 2006. 18th International Conference.

Pentland, A. 2008. Honest Signals: how they shape our world. MIT Press; Cambridge, MA.

Buchanan, M. 2009. Secret signals: Does a primitive, non-linguistic type of communication drive people's interactions? Nature 457:528–530.

Pentland, A (2010). To Signal Is Human. American Scientist 98(3) May-June 2010 203-211.

Pentland, A (2011): "Signals and speech", In *INTERSPEECH-2011*, 1-4.

Schlangen, D. From reaction to prediction: Experiments with computational models of turn-taking. In Ninth International Conference on Spoken Language Processing. Citeseer, 2006.

Vinciarelli, A , Pantic ,M, Bourlard, and Pentland, A. 2008. Social signals, their function, and automatic analysis: a survey. In Proceedings of the 10th international conference on Multimodal interfaces (ICMI '08). ACM, New York, NY, USA

Miscellaneous

Brand M, Oliver N, and Pentland A. 1997. Coupled hidden Markov models for complex action recognition. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 994–999.

Chen, K-Y, Fine, L., and Huberman, B., “Predicting the Future,” Information Systems Frontiers, Vol. 5, No. 1, pp. 47-61, 2004

Jordan MI, Ghahramani Z, Jaakkola TS, and Saul LK. 1999. An introduction to variational methods for graphical models. Machine learning, 37(2):183–233.

Pentland A, Eagle N. 2009. Eigenbehaviors: Identifying structure in routine Behavioral Ecology and Sociobiology 63 (7).

Technical Area: Multimodal Grounded Learning

Examiner

Prof. Patrick Henry Winston
Professor of Electrical Engineering & Computer Science
MIT Computer Science and Artificial Intelligence Lab

Description

The technical area of this exam provides a strong technical foundation for studying grounded semantics. This area will reviews a wide range of topics such as grounded and multi-modal language learning, established (and new) work in language processing techniques (with emphasis on semantics) and perception and knowledge representation. In addition, this area will cover foundational readings on computational, cognitive (and to a lesser extent linguist) theories behind these concepts.

Requirement

The written requirement for this area will be a 24 hour take home examination to be administered and evaluated by Prof. Patrick Winston.

Examiner's signature: Date:

Reading List

Grounded Language Learning

Bailey, D. 1997. When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs. Ph.D. Dissertation

Branavan, S. R. K.; Chen, H.; Zettlemoyer, L. S.; and Barzilay, R. 2009. Reinforcement learning for mapping instructions to actions. In Proceedings of ACL, 82–90.

Bugmann, G.; Klein, E.; Lauria, S.; and Kyriacou, T. 2004. Corpus-based robotics: A route instruction example. Proceedings of Intelligent Autonomous Systems 96—103

Gold, K., & Scassellati, B. (2007). A robot that uses existing vocabulary to infer non-visual word meanings from observation.

Kollar, T.; Tellex, S.; Roy, D.; and Roy, N. 2010a. Grounding verbs of motion in natural language commands to robots. In Proceedings of the International Symposium on Experimental Robotics (ISER)

Marocco, D.; Cangelosi, A.; Fischer, K.; and Belpaeme, T. 2010. Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. Frontiers in Neurorobotics 4(0)

Modayil, J., and Kuipers, B. 2007. Autonomous development of a grounded object ontology by a learning robot. In Proc. AAAI, volume 2, 1095–1101. AAAI Pres

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition, 61, 39-91.

Tellex, S.; Kollar, T.; Shaw, G.; Roy, N.; and Roy, D. 2010. Grounding spatial language for video search. In International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI '10, 31:1–31:8. ACM

Tellex S, Kollar T, Dickerson S, Walter MR, Banerjee AG, Teller S and Roy N. (2011) . Approaching the Symbol Grounding Problem with Probabilistic Graphical Models. AI Magazine Special Issue on Dialogue with Robots (Vol. 32, No. 4), pp. 64-76.

Knowledge Representation

Dretske, F. I. Knowledge and the flow of information. CSLI publications; Stanford CA: 1999.

Dzifcak, J.; Scheutz, M.; Baral, C.; and Schermerhorn, P. 2009. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. In Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA), 4163–4168

G. C. Borchardt. A computer model for the representation and identification of physical events. Master's thesis, report t-142, University of Illinois, Coordinated Science Laboratory, University of Illinois, Urbana,. IL, 198

G. C. Borchardt. Casual reconstruction. A.I. Memo 1403, Massachusetts Institute of Technology, Aritificial Intelligence Laboratory, Cambridge, Massachusetts, February 1993

Winograd, T. 1970. Procedures as a representation for data in a computer program for understanding natural language. Ph.D. Dissertation, Massachusetts Institute of Technology

Perception, Simulation and Analogy

Gentner D, Holyoak K.J, and Kokinov N, The Analogical Mind: Perspectives from Cognitive Science. Cambridge, MA:MIT Press, 2001.

Holyoak K.J, and Thagard P. Mental Leaps : Analogy in Creative Thought. A Bradford Book, the MIT Press, Cambridge, Massachusetts, 1995.

Hsiao, K.-y.; Mavridis, N.; and Roy, D. 2003. Coupling perception and simulation: Steps towards conversational robotics. In Proceedings of the IEEE/RSJ International.

Smith BC. 1998. On the Origin of Objects. Cambridge, MA: MIT Press

Winston, PH. The Strong Story Hypothesis and the Directed Perception Hypothesis. Advances in Cognitive Systems, AAAI Fall Symposium, 2001.

Language Processing

de Marneffe, M.; MacCartney, B.; and Manning, C. 2006. Generating typed dependency parses from phrase structure parses. In Proc. Int'l Conf. on Language Resources and Evaluation (LREC), 449–454.

Ge, R., and Mooney, R. J. 2005. A statistical semantic parser that integrates syntax and semantics. In Proc. of the Ninth Conference on Computational Natural Language Learning, 9–16. Stroudsburg, PA, USA: Association for Computational Linguistics

Jones, B., Johnson, M., & Frank, M. C. (2010). Learning words and their meanings from unsegmented child-directed speech. In Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Fazly, A., Alishahi, A., & Stevenson, S. (in press). A probabilistic computational model of cross-situational word learning. Cognitive Science.

Langley, P and Carbonell, GJ. Language acquisition and machine learning. In Brian MacWhinney, editor, Mechanisms of Language Acquisition, chapter 5, pages 115–155. Lawrence Erlbaum Associates, 1987.

Marcken, CD. Unsupervised Language Acquisition. PhD thesis, Massachusetts Institute of Technology, September 1996.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. Journal of Memory and Language, 39(246-263).

Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. Natural Language Engineering, 3(01), 57-87.

Regier, T. P. 1992. The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization. Ph.D. Dissertation, University of California at Berkeley

Stickgold, E. (2011). Word Sense Disambiguation Through Lattice Learning. Unpublished master's thesis, Massachusetts Institute of Technology.

Theory and Techniques

Barwise, J. & Perry, J. (1983) Situations and attitudes. MIT Press.

Chapman, D. (1992). Vision, instruction and action. Cambridge, MA: The MIT Press.

Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.

Correa, T L. (2011). A Model for Transition-Based Visuospatial Pattern Recognition. Unpublished master's thesis, Massachusetts Institute of Technology.

Jackendoff, R. S. 1983. Semantics and Cognition. MIT Press. 161–187.

Lieberman, P. (2002). Human language and our reptilian brain: the subcortical bases of speech, syntax, and thought. Cambridge, MA: Harvard University Press.

McClelland, J. L., and Rumelhart, D. E. 1989. Explorations in Parallel Distributed Processing.

Pierce, JR. An Introduction to Information Theory: Symbols, Signals & Noise. Dover Publications, 1980.

Satyajit Rao. Visual Routines and Attention. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, February 1998

Schank RC, Abelson RP (1977). Goals, and Understanding: An Inquiry into Human Knowledge Structures. Lawrence Elbraum.

Sugita, Y., and Tani, J. 2005. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems 13:33–52

Winston PH. Learning Structural Descriptions From Examples. PhD thesis, Massachusetts Institute of Technology, 1970

Winston PH. Learning by analyzing differences. In Articial intelligence (3. ed.), pages 349-363. Addison-Wesley, 1992. ISBN 978-0-201-53377-4.