

**LT2318: Artificial Intelligence: Cognitive Systems (AICS)**  
**Reading list**

Simon Dobnik  
Centre for Linguistic Theory and Studies in Probability (CLASP)  
University of Gothenburg  
simon.dobnik@gu.se

October 18, 2018

- Agrawal, A., D. Batra, D. Parikh, and A. Kembhavi (2017). Don't just assume; look and answer: Overcoming priors for visual question answering. *arXiv arXiv:1712.00377 [cs.CV]*.
- Akram, S. U. (2012). Visual recognition of isolated swedish sign language signs. Master's thesis, School of Computer Science and Communication, Control and Robotics, Royal Institute of Technology, Stockholm, Sweden.
- Andreas, J., M. Rohrbach, T. Darrell, and D. Klein (2016, June 12-17). Learning to compose neural networks for question answering. In *Proceedings of NAACL-HLT 2016*, San Diego, California, pp. 1545–1554. Association for Computational Linguistics.
- Åstbom, A. (2017, February 7). How function of objects affects geometry of spatial descriptions. A study of Swedish and Japanese. C-uppsats (bachelor's thesis/extended essay), Department of Philosophy, Linguistics and Theory of Science (FLOV), University of Gothenburg, Gothenburg, Sweden. Supervisor: Simon Dobnik, opponent: Linnea Strand, examiner: Christine Howes.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–609.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology* 59, 617–645.
- Battaglia, P. W., J. B. Hamrick, and J. B. Tenenbaum (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences* 110(45), 18327–18332.
- Bradski, G. and A. Kaehler (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.
- Bruni, E., N.-K. Tran, and M. Baroni (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research (JAIR)* 49(1–47).
- Byron, D. K. (2003). Understanding referring expressions in situated language some challenges for real-world agents. In *Proceedings of the First International Workshop on Language Understanding and Agents for Real World Interaction*, pp. 39–47.
- Cooper, R. (in prep). Type theory and language: from perception to linguistic communication. Draft of book chapters available from <https://sites.google.com/site/typetheorywithrecords/drafts>.
- Coventry, K. and S. Garrod (2005). Spatial prepositions and the functional geometric framework. towards a classification of extra-geometric influences. In L. A. Carlson and E. v. d. Zee (Eds.), *Functional features in language and space: insights from perception, categorization, and development*, Volume 2, pp. 149–162. Oxford University Press.

- Coventry, K. R., A. Cangelosi, R. Rajapakse, A. Bacon, S. Newstead, D. Joyce, and L. V. Richards (2005). Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky (Eds.), *Spatial Cognition IV. Reasoning, Action, Interaction*, Volume 3343 of *Lecture Notes in Computer Science*, pp. 98–110. Springer Berlin Heidelberg.
- de Graaf, E. (2016, June, 8th). Learning objects and spatial relations with Kinect. Master's thesis, Department of Philosophy, Linguistics and Theory of Science. University of Gothenburg, Gothenburg, Sweden. Supervisor: Simon Dobnik, examiner: Richard Johansson, opponent: Lorena Llozhi.
- Dissanayake, M. W. M. G., P. M. Newman, H. F. Durrant-Whyte, S. Clark, and M. Csorba (2001). A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotot and Automation* 17(3), 229–241.
- Dobnik, S. (2009, September 4). *Teaching mobile robots to use spatial words*. Ph. D. thesis, University of Oxford: Faculty of Linguistics, Philology and Phonetics and The Queen's College, Oxford, United Kingdom. <http://www.dobnik.net/simon/documents/thesis.pdf>.
- Dobnik, S. and A. Åstbom (2017, August 15–17). (Perceptual) grounding as interaction. In V. Petukhova and Y. Tian (Eds.), *Proceedings of Saardial – Semdial 2017: The 21st Workshop on the Semantics and Pragmatics of Dialogue*, Saarbrücken, Germany, pp. 17–26.
- Dobnik, S. and E. de Graaf (2017, 22–24 May). KILLE: a framework for situated agents for learning language through interaction. In J. Tiedemann and N. Tahmasebi (Eds.), *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, Gothenburg, Sweden, pp. 162–171. Northern European Association for Language Technology (NEALT): Association for Computational Linguistics.
- Dobnik, S., C. Howes, and J. D. Kelleher (2015, 24–26th August). Changing perspective: Local alignment of reference frames in dialogue. In C. Howes and S. Larsson (Eds.), *Proceedings of goDIAL – Semdial 2015: The 19th Workshop on the Semantics and Pragmatics of Dialogue*, Gothenburg, Sweden, pp. 24–32.
- Dobnik, S. and J. D. Kelleher (2013, 31 July). Towards an automatic identification of functional and geometric spatial prepositions. In *Proceedings of PRE-CogSsci 2013: Production of referring expressions – bridging the gap between cognitive and computational approaches to reference*, Berlin, Germany, pp. 1–6.
- Dobnik, S. and J. D. Kelleher (2014, August). Exploration of functional semantics of prepositions from corpora of descriptions of visual scenes. In *Proceedings of the Third V&L Net Workshop on Vision and Language*, Dublin, Ireland, pp. 33–37. Dublin City University and the Association for Computational Linguistics.
- Dobnik, S., J. D. Kelleher, and C. Koniaris (2014, 1–3 September). Priming and alignment of frame of reference in situated conversation. In V. Rieser and P. Muller (Eds.), *Proceedings of DialWatt – Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue*, Edinburgh, pp. 43–52.
- Foerster, J. N., Y. M. Assael, N. de Freitas, and S. Whiteson (2016). Learning to communicate with deep multi-agent reinforcement learning. *CoRR abs/1605.06676*.
- Ghanimifard, M. and S. Dobnik (2017, September 19–22). Learning to compose spatial relations with grounded neural language models. In C. Gardent and C. Retoré (Eds.), *Proceedings of IWCS 2017: 12th International Conference on Computational Semantics*, Montpellier, France, pp. 1–12. Association for Computational Linguistics.

- Glenberg, A. M. (1997, April). What memory is for. *The Behavioral and brain sciences* 20, 1–55.
- Goebel, P. (2013). *ROS by example*. Lulu.
- Hamilton, A. F. d. C., K. Kessler, and S. H. Creem-Regehr (2014, June). Perspective taking: building a neurocognitive framework for integrating the “social” and the “spatial”. *Frontiers in Human Neuroscience* 8(403), 1–3.
- Harnad, S. (1990, June). The symbol grounding problem. *Physica D* 42(1–3), 335–346.
- Jordan, P. W. and M. A. Walker (2005). Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research* 24, 157–194.
- Karpathy, A. and L. Fei-Fei (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137.
- Kelleher, J. D. and F. J. Costello (2009). Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics* 35(2), 271–306.
- Kelleher, J. D., F. J. Costello, and J. van Genabith (2005). Dynamically structuring updating and interrelating representations of visual and linguistic discourse. *Artificial Intelligence* 167, 62–102.
- Kennington, C., S. Kousidis, and D. Schlangen (2014). InproTKs: a toolkit for incremental situated processing. *Proceedings of SIGdial 2014: Short Papers*.
- Kennington, C. and D. Schlangen (2015, July). Simple learning and compositional application of perceptually grounded word meanings for incremental reference resolution. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China, pp. 292–301. Association for Computational Linguistics.
- Krafka, K., A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba (2016). Eye tracking for everyone. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Krahmer, E. and K. van Deemter (2011). Computational generation of referring expressions: A survey. *Computational Linguistics* 38(1), 173–218.
- Kruijff, G.-J. M., H. Zender, P. Jensfelt, and H. I. Christensen (2007). Situated dialogue and spatial organization: what, where... and why? *International Journal of Advanced Robotic Systems* 4(1), 125–138. Special issue on human and robot interactive communication.
- Lang, S. (2011). Sign language recognition with kinect. Bachelorarbeit, Institut für Informatik, Freie Universität Berlin, Berlin, Germany.
- Lang, S., M. Block, and R. Rojas (2012). *Sign Language Recognition Using Kinect*, pp. 394–402. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lauria, S., G. Bugmann, T. Kyriacou, J. Bos, and E. Klein (2001, September/October). Training personal robots using natural language instruction. *IEEE Intelligent Systems* 16, 38–45.
- Lauria, S., G. Bugmann, T. Kyriacou, and E. Klein (2002). Mobile robot programming using natural language. *Robotics and Autonomous Systems* 38(3–4), 171–181.
- Lazaridou, A., A. Peysakhovich, and M. Baroni (2016). Multi-agent cooperation and the emergence of (natural) language. *arXiv arXiv:1612.07182v2 [cs.CL]*, 1–11.
- Logan, G. D. and D. D. Sadler (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett (Eds.), *Language and Space*, pp. 493–530. Cambridge, MA: MIT Press.

- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Volume 2, pp. 1150–1157. IEEE.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), 91–110.
- Lu, J., C. Xiong, D. Parikh, and R. Socher (2017, 6 June). Knowing when to look: Adaptive attention via a visual sentinel for image captioning. arXiv:1612.01887 [cs.CV].
- Malinowski, M., M. Rohrbach, and M. Fritz (2015). Ask your neurons: A neural-based approach to answering questions about images. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1–9.
- Marr, D. (2010, August). *Vision: A computational approach*. MIT Press Scholarship Online.
- Matuszek, C., N. FitzGerald, L. Zettlemoyer, L. Bo, and D. Fox (2012, June 27th - July 3rd). A joint model of language and perception for grounded attribute learning. In J. Langford and J. Pineau (Eds.), *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, Edinburgh, Scotland.
- Matuszek, C., E. Herbst, L. Zettlemoyer, and D. Fox (2012, June). Learning to parse natural language commands to a robot control system. In *Proceedings of the 13th International Symposium on Experimental Robotics (ISER)*.
- McMahan, B. and M. Stone (2015). A Bayesian model of grounded color semantics. *Transactions of the Association for Computational Linguistics* 3, 103–115.
- Mei, H., M. Bansal, and M. R. Walter (2016). Listen, attend, and walk: Neural mapping of navigational instructions to action sequences. In *AAAI*, pp. 2772–2778.
- Monroe, W., N. D. Goodman, and C. Potts (2016). Learning to generate compositional color descriptions. *CoRR abs/1606.03821*.
- Muja, M. and D. G. Lowe (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)* 2(331–340), 2.
- Mustafa, E. and K. Dimopoulos (2014). Sign language recognition using kinect. In *Dautov, R., Gkasis, P., & Karama-nos, A. et al.(2014). Proceedings of the 9th South East European Doctoral Student Conference*, pp. 271–285. Thessaloniki: SEERC.
- O’Kane, J. M. (2013, 10). *A Gentle Introduction to ROS*. CreateSpace Independent Publishing Platform.
- Pezzelle, S., I. Sorodoc, and R. Bernardi (2018). Comparatives, quantifiers, proportions: A multi-task model for the learning of quantities from vision. *arXiv arXiv:1804.05018 [cs.CV]*, 1–12.
- Quigley, M., K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng (2009). ROS: an open-source robot operating system. In *ICRA workshop on open source software*, Volume 3, pp. 5.
- Rashtchian, C., P. Young, M. Hodosh, and J. Hockenmaier (2010, 6 June). Collecting image annotations using Amazon’s Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on creating speech and language data with Amazon’s Mechanical Turk*, Los Angeles, CA. North American Chapter of the Association for Computational Linguistics (NAACL).
- Regier, T. (1996). *The human semantic potential: spatial language and constrained connectionism*. Cambridge, Massachusetts, London, England: MIT Press.
- Regier, T. and L. A. Carlson (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General* 130(2), 273–298.

- Roy, D. (2002). Learning visually-grounded words and syntax for a scene description task. *Computer speech and language* 16(3), 353–385.
- Roy, D. (2005, September). Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence* 167(1-2), 170–205.
- Skantze, G. (2016). Real-time coordination in human-robot interaction using face and voice. *AI Magazine* 37(4), 19–31.
- Skantze, G. and S. Al Moubayed (2012). IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pp. 69–76. ACM.
- Skantze, G., A. Hjalmarsson, and C. Oertel (2014). Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication* 65, 50–66.
- Skočaj, D., M. Janiček, M. Kristan, G.-J. M. Kruijff, A. Leonardis, P. Lison, A. Vrečko, and M. Zillich (2010). A basic cognitive system for interactive continuous learning of visual concepts. In *ICRA 2010 workshop ICAIR - Interactive Communication for Autonomous Intelligent Robots*, Anchorage, AK, USA, pp. 30–36.
- Skočaj, D., M. Kristan, A. Vrečko, M. Mahnič, M. Janiček, G.-J. M. Kruijff, M. Hanheide, N. Hawes, T. Keller, M. Zillich, and K. Zhou (2011, 25-30 September). A system for interactive learning in dialogue with a tutor. In *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS 2011*, San Francisco, CA, USA.
- Steels, L. and J.-C. Baillie (2003, May). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems* 43(2–3), 163–173.
- Steels, L. and T. Belpaeme (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences* 28(4), 469–489.
- Steels, L. and M. Loetzsch (2009). Perspective alignment in spatial language. In K. R. Coventry, T. Tenbrink, and J. A. Bateman (Eds.), *Spatial Language and Dialogue*. Oxford University Press.
- Stoia, L., D. M. Shockley, D. K. Byron, and E. Fosler-Lussier (2006, July). Noun phrase generation for situated dialogs. In *Proceedings of the Fourth International Natural Language Generation Conference*, Sydney, Australia, pp. 81–88. Association for Computational Linguistics.
- Tenenbaum, J. B., C. Kemp, T. L. Griffiths, and N. D. Goodman (2011). How to grow a mind: Statistics, structure, and abstraction. *Science* 331(6022), 1279–1285.
- Viethen, J., R. Dale, and M. Guhe (2011a). Generating subsequent reference in shared visual scenes: Computation vs. re-use. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1158–1167. Association for Computational Linguistics.
- Viethen, J., R. Dale, and M. Guhe (2011b). The impact of visual context on the content of referring expressions. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pp. 44–52. Association for Computational Linguistics.
- Wang, S. I., P. Liang, and C. D. Manning (2016). Learning language games through interaction. *arXiv preprint arXiv:1606.02447*.
- Winograd, T. (1972). Understanding natural language. *Cognitive Psychology* 3(1).
- Winograd, T. (1976). *Understanding Natural Language*. Edinburgh University Press.

- Xu, K., J. Ba, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio (2015, February 11). Show, attend and tell: Neural image caption generation with visual attention. In *arXiv preprint*, Number arXiv:1502.03044v3 [cs.LG], pp. 1–22.
- Yang, H.-D. (2014). Sign language recognition with the kinect sensor based on conditional random fields. *Sensors* 15(1), 135–147.
- Zender, H., M. Janicek, and G.-J. Kruijff (2012). Situated communication for joint activity in human-robot teams. *IEEE Intelligent Systems* 27(2), 27–35.
- Zender, H., O. Martínez-Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard (2008, June). Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems* 56(6), 493–502. Special issue “From sensors to human spatial concepts”.