Research Statement
Cynthia Matuszek

Research advances in robotics and other cyberphysical systems are enabling more and more sophisticated, capable technologies to reach large consumer populations. However, natural, easy-to-use interfaces to such systems lag far behind. Such systems offer unprecedented opportunities in a variety of human-centric applications such as elder care, assistive care, and household maintenance; as robots become more capable and ubiquitous in these settings, the importance of allowing non-expert users to interact with them naturally—using communication modalities with which people are comfortable—increases. Reaching this goal will require technical advances in a number of areas, including robotics, natural language processing, machine learning, and human-computer-interaction.

My research aims to enable robots to usefully participate in everyday human activity and learn to understand commands expressed in human language, using gesture and other forms of interaction as guides. This work is at the intersection of robotics and language processing, with a focus on building useful robotic systems that can be deployed in shared spaces with non-expert users. Such systems must accomplish a number of goals: they must understand the user’s (possibly multimodal) communications; they must learn about objects in the world and tasks to be performed, using a combination of language and sensor data; and they must perform those tasks satisfactorily and safely, without inducing unnecessary frustration or cognitive load in users.

My work to date has focused on building robotic systems that can work safely in shared spaces with humans, and can learn, from unscripted interactions, how to follow commands and understand language. This is a fundamentally interdisciplinary task, and an important aspect of my approach to research is to work with collaborators with diverse expertise, including knowledge representation, computer vision, and security, as well as natural language processing and human-robot interaction. While creating usable robotic systems is a unifying application, the advances I make in each of the sub-problems will be applicable to many other research challenges in natural language processing, artificial intelligence, and cyberphysical systems.

Current Research

My current research focuses on learning novel concepts on the fly, from language and interaction. I have worked on (1) grounded language acquisition: learning to understand novel concepts, by extracting semantically meaningful representations of language and mapping those representations to the noisy, unpredictable physical world in which robots operate; (2) learned instruction following: using learned semantic understanding to convert natural language commands into robot control code, so the system can learn to perform tasks; (3) multimodal input understanding: combining language learning with learning from examples of how people use gestures to direct attention and instruct a robot; and (4) a shared-space robotic platform: the development and use of a manipulation platform which interacts successfully and safely in shared spaces with end users. I have demonstrated the use of this platform in game-playing and in acting on concepts learned from interaction with people.

Learning to Interpret Language about Objects. Advances in both natural language and physical agents have led to intense study of the field of natural language grounding, or interpreting language in the context of sensing and actuation. I study the application of semantic parsing – parsing natural language into a formal representation of its intended meaning – to building systems that understand commands and descriptions of the world. The key idea is to train a semantic parser with data drawn from a combination of language and sensor input; underlying linguistic meaning is learned from real-world examples. These underlying meanings are expressed using a probabilistic version of the Combinatory Categorial Grammar (CCG) formalism, a type of phrase structure grammar. CCGs model both the syntax (language constructs, such as NP for noun phrase) and the semantics (expressions in λ-calculus) of a sentence.

To perform grounded language learning, a parser is trained by inducing a probabilistic CCG (PCCG) from a set of training examples that include both words and percepts from a sensor. By connecting symbols found in language to novel sensor inputs, the meanings of new words can be induced. This entails
reasoning jointly about language and vision. Given a simple sentence such as “These are the yellow cubes,” uttered in a setting where there is a physical workspace that contains a number of objects that vary in shape and color, the robot must realize that words such as “yellow” and “cubes” refer to object attributes, and ground these words by mapping them to a perceptual system that will enable it to identify the specific physical objects being referred to. Training such a parser yields a robot that can learn, from natural language and input from a camera, how to interpret unconstrained natural language describing physical characteristics of objects [2]. This system can be taught about objects in the world and how they are referred to, by its end-user, without a programmer needing to consider what those objects might be and without being re-programmed as the user’s needs change.

Parsing Language Into a Robot Control System. It’s generally possible to hand-engineer systems to accomplish specific, well-defined tasks. However, for a robot in an unconstrained human environment, the variety of possible applications and the complexity of physical settings mean that learning must be a core element of a flexible, useful system. I introduced an approach for learning the meanings of instructional language (used to give directions to a mobile robot in simulation), using two different logical formalisms: PCCGs as described above [5], and probabilistic synchronous context-free grammars, which are trained using techniques from statistical machine translation to learn probabilities of bidirectional derivation rules between natural language and a logical formalism [3]. In this work, I collected English route instructions describing a path through a map, and a formal description of the path was generated automatically or semi-automatically [3, 5]. Pairs of natural language and formal representation are used as training inputs. The learned parser was then tasked with interpreting novel route instructions in a novel map, and we evaluated a simulated agent’s ability to follow those instructions successfully.

The formal grammar developed for representing route instructions in this work is able to capture English concepts that have complex procedural interpretations when applied to robot control; the compositional character of parser learning allowed for generalization of concepts, so that concepts such as counting (“take the third door”) or analogous reasoning (“take a right” versus “take a left”) need not be learned individually. As a result, unlike previous work, the resulting system learns from a relatively small amount of English training data, and can perform simultaneous exploration of a novel map while following instructions through it. Analogous to learning about objects in the world, in this work a robot learns how to follow commands from language and examples, rather than performing fully scripted tasks.

Human-Robot Interaction using Language and Gesture. While learning to understand language is an important element of HRI, people naturally communicate using rich, multimodal interactions. Language commands about physically situated topics can be tortuous without physical attentional cues such as pointing gestures (for example, compare “Put the mug that’s on the table in the cupboard to the left of the stove,” versus “Put this mug in there”). In addition, speech is a naturally noisy signal for communication – spoken language is full of non-grammatical or disfluent statements, and speech recognition introduces additional sources of error. Other communication modes such as gesture provide additional learning signal, and support understanding natural, frustration-free communication.

To take advantage of this signal, I investigated how people refer to objects in the world during relatively unstructured communication with robots. A corpus of unscripted interactions with non-experts was collected (Fig. 1) and used to train novel models of language and gesture. In order to successfully train a classifier based on the varied attentional gestures in this corpus, we developed a temporal extension to state-of-the-art hierarchical matching pursuit features. This classifier was combined with a language model learned from speech in the corpus, and implemented on an instruction-following robot (Fig. 2). This system learns to understand language, as above, but also goes beyond language to engage with a person in a physical setting, using sensors to understand gestural communication and actuation to follow simple commands using that combination of modalities [1]. (Video: http://tiny.cc/Gambit14)

Robotic Platform and Game Playing in Shared Spaces. In order to demonstrate that language acquisition and gesture understanding have the potential to be useful in human-robot interaction, it is necessary to have a robotic platform that can be used safely for interactive tasks. I was part of the effort to design and build the Gambit small-scale manipulation platform, which was developed in collaboration among Intel, Allium Labs, and the University of Washington. The robot was designed to perform tabletop-scale tasks with
high precision at relatively low cost. This platform was used successfully both to engage in relatively unstructured gameplay and to follow unscripted human instructions in a variety of settings, demonstrating the potential utility of these approaches to human-robot interaction.

The platform’s initial test was the AAAI 2010 small-scale manipulation challenge, in which different robots played chess against each other and human intermediaries, as well as against human challengers. Physical board games are a rich example problem domain for human-robot cooperation research, involving perception of the world, perception of human action, and careful manipulation, while coordinating with a possibly unpredictable human opponent. Our system interacts smoothly with opponents who are given no special instructions or training (Fig. 3), using arbitrary chess sets on a variety of boards, and requiring no instrumentation or modeling of pieces. The opponent's moves are detected, and the perceptual system tracks the board position in real time. Our system achieved the top score at the AAAI challenge. (Video: http://tiny.cc/Chess11)

Future Directions

Human-Robot Interaction spans a wide range of topics and presents the challenge of building integrated systems that combine the advances from each area. Ultimately, I want to build systems that allow users to interact naturally with a variety of small robots that can be deployed in traditionally human-centric settings. Such robots will be deployed with some understanding of objects in the world and their own duties, but will also be capable of learning from their users, guided by not only language and gesture, but behavior and affect. Central to this goal is the creation and testing of human-robot interfaces that learn efficiently and interact naturally. To that end, I plan to investigate:

**Human-Robot Dialog for Active Learning.** To go beyond learning passively by listening to a human teacher, it is necessary for a robot to ask questions or otherwise help direct its own learning; such active learning allows for more efficient learning and better performance as well as more natural interaction. Using active learning in human-robot interactions presents its own challenges, raising issues not only of what questions to ask, but when and how to ask them: We must balance the improvement in the robot’s acquisition of concepts with the need not to make excessive demands of users. The goal is then to incorporate only as much questioning as is useful to drive efficient understanding of human utterances.

Much of the work on active learning in robotics settings revolves around dialog systems, and, in particular, around an architecture in which a robot has a goal or goals that are integrated into a central architecture that may include other goals such as accomplishing tasks. I intend to explore the utility of making a robot that uses language acquisition to be a more active participant in its own learning. This includes taking advantage of the robot’s ability to incorporate nonverbal elements of queries (such as pointing to an object while asking about it, or using tone to convey uncertainty), leading to more transparent and therefore hopefully more natural interactions.

**Additional Learning Mechanisms to Supplement Human Teaching.** Robots should be able to learn new concepts from their users, but they should be supplemental to other sources of information. I intend to follow up on previous work on automatically extending formal representations of the world [7] in order to help robots gather information about new concepts, especially in circumstances where the transaction cost
of asking a user for help is too high, or simply unnecessary. For example, if a user asks a robot to retrieve an unfamiliar object such as a Coke can from another room, an online image search provides enough information to attempt a solution without asking for additional help.

Crowdsourcing, in particular, has the potential to address a number of otherwise cumbersome problems. Large datasets can be gathered in order to provide a solid basis for language learning [1], but more focused uses will allow deployed robots to solve specific problems. Pre-emptive information gathering, such as asking for labels when an unrecognized object is in view, can be continuously conducted; higher-cost, real-time crowdsourcing can help with time-critical problems such as finding an elevator button in a camera view. The addition of these mechanisms to existing systems will allow deployed robots to continuously improve performance without constantly requesting user assistance.

**Assistive Technology.** One of the most immediate and crucial areas in which robots can be deployed productively is in the developing areas of assistive technology and elder care. Systems that can supplement human capabilities – such as by providing descriptions of the environment to visually impaired persons [8] or performing shared manipulations with a one-handed user – have the potential to be useful immediately, setting a lower bar for the technology and offering the potential for immediate value even when the deployed systems are limited. Robots that increase autonomy for populations such as aging-in-place elders have the potential to have broad impact and attract substantial funding, but also require careful user studies, collaboration with entities outside of computer science, and sensitivity to the costs and benefits of the systems being developed. I intend to collaborate with both human-computer interaction specialists and colleagues in appropriate external groups, such as medical research, psychology, and human-centered design.

**Improved Safety of Physical Systems.** As the goals described here are achieved, we will begin to see a proliferation of special-purpose robots in homes, workspaces, and other human-populated settings. Current systems often operate in carefully constrained situations, with dedicated “robot only” workspaces and highly trained users. For robots to move into shared spaces with people, safety and security need to be considered carefully; however, our study [6] of the security of small household robots suggests that current reality falls short of even basic security best practices (for example, by failing to giving telepresence robots instructions about where to go in a home over an encrypted channel). One approach to providing guarantees of safe behavior is to verify the correctness of commands – such as those from our semantic parser – against a formally represented set of requirements. I plan to apply such formal verification approaches to commands a robot takes from a human, in order to provide behavior guarantees. I also intend to seek continued collaboration with security and cyberphysical systems experts.

**Summary**

My research to date has focused on combining robotics and natural language processing to both drive natural language learning and to improve robotics, by building natural, human-centric interfaces for robots that can interact with non-specialist users. My future research will be focused on the multidisciplinary challenges of creating systems that will be useful in the near term. I believe my research will attract funding from both academic funding agencies and industry. My research aligns well with the NSF’s National Robotics Initiative, which has an emphasis on both collaborative robotics and assistive care. My academic studies are co-advised between two research groups, and I have additionally collaborated on papers with security and HCI researchers, leaving me well positioned to collaborate with colleagues in a variety of areas. Ultimately, I hope my work will contribute to deployed, safe robotics that are easy for any population to interact with and control.
References


Teaching Statement
Cynthia Matuszek

Teaching is an important factor in my decision to pursue an academic career. At its best, good teaching provides students with tools for the future, supplies well-educated and engaged future members of the field, and gives the teacher with a clear, well-organized mental overview of a subject, as well as tremendous personal satisfaction. In computer science, we have two advantages: first, the opportunity to showcase the range of interesting problems to which computers can be applied; second, the ability to develop and use novel technology to mentor students whose interests or needs go beyond the traditional classroom setting. We are uniquely positioned to use technology to support and drive education.

Previous Experience
I assisted the teaching of two undergraduate level courses, both in Artificial Intelligence. My role in these classes involved preparing and teaching a subset of the lectures, grading assignments and exams, and holding office and lab hours. I have also given guest lectures in graduate and undergraduate computer science classes. I find lecturing the most challenging and rewarding. Keeping students engaged and entertained while reacting appropriately to their ongoing level of comprehension requires both preparation and flexibility.

Mentoring and Outreach
I have supervised a number of undergraduate and early Ph.D. students in their research. The most successful collaborations were those in which we identified challenging problems that corresponded well to the interests, abilities, and time commitments of the students, but still advanced the needs of a broader research agenda; in particular, this is true for students that I mentored as part of a research project in which they had ownership of well-defined areas, such as integrating a chess problem-solver into a robot chess player. Two of these collaborations led to student co-authorships on conference publications.

Besides the core responsibilities of teaching and mentoring, I am dedicated to outreach and increasing the diversity of the student population in computer science, including reaching out to women and minority populations. I have been invited to demonstrate my research to younger and non-technical audiences on many occasions, and have consistently gotten positive, engaged responses. Examples include giving presentations at the summer academy for advancing deaf and hard of hearing in computing and to students from the AVID (Advancement Via Individual Determination) program for underserved populations. Engaging students with a variety of backgrounds and needs is a deeply rewarding challenge, and I intend to continue these activities throughout my academic career.

Teaching Interests
I am interested in teaching in a range of courses. Possible introductory courses include basic programming and data structures; advanced courses include robotics, artificial intelligence, natural language processing, and, if desired, human-computer interaction. My goals in teaching such classes are twofold. First, to present material in a way that makes it clear how each class fits into the overarching curriculum and field, and how topics are mutually relevant rather than disconnected – for example, teaching students about reinforcement learning while showing them a robot that is learning to walk, or describing elements of HCI by giving examples of developing games versus IDEs. Second, I work to make classes as engaging and effective as possible by making the best possible use of new supporting technologies: for example, large-scale lecture distribution and telepresence to reach additional populations, and social media and real-time class feedback devices to improve engagement.