2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Lingodroids: Studies in Spatial Cognition and Language

Ruth Schulz, Arren Glover, Michael J. Milford, *Member, IEEE*, Gordon Wyeth *Member, IEEE*, Janet Wiles

Abstract— **The Lingodroids are a pair of mobile robots that evolve a language for places and relationships between places (based on distance and direction). Each robot in these studies has its own understanding of the layout of the world, based on its unique experiences and exploration of the environment. Despite having different internal representations of the world, the robots are able to develop a common lexicon for places, and then use simple sentences to explain and understand relationships between places – even places that they could not physically experience, such as areas behind closed doors. By learning the language, the robots are able to develop representations for places that are inaccessible to them, and later, when the doors are opened, use those representations to perform goal-directed behavior.**

I. Introduction

ROBOTS are a powerful tool for the study of cognitive processes. In this paper, we illustrate interactions between two robots that model cognitive processes ranging from knowledge representation and planning to language development, symbol grounding and even imagination. The studies we present are based on two real robots, called Lingodroids, that autonomously develop a sophisticated language to negotiate spatial tasks in, and beyond, their individual cognitive maps.

Cognitive maps are a class of biologically-inspired maps that ground spatial knowledge in the sensing and behavior of the robot, rather than defining a map by the location of features in a geometric frame. Cognitive maps have been shown to be powerful tools for challenging robotic problems such as visual mapping and localization [1], and long term mapping and navigation [2]. The models that produce cognitive maps provide valuable insight into the functional properties of the mammalian brain [3, 4].

Consider two robots that independently build cognitive maps of an area, each grounding experience of the area through its own sequence of actions and sensor readings. If these maps are effective, each robot should be able to use its own map to navigate to a goal location. But how could one robot use the cognitive map of another robot to perform navigation tasks? How can the two robots share knowledge that has been assimilated into two heterogeneous representations of the same space? For effective communication, the two robots need a system of shared symbols, where the symbols have the same meaning to each robot. In cognitive science terms, the robots must solve the *symbol grounding problem* [5].

One solution to the symbol grounding problem that has been shown to be effective in robotic systems is the system for autonomous acquisition of language called *language games* [6]. In a language game, two agents choose a common topic by a mechanism of shared attention. One agent generates an utterance to describe the subject at hand, while the other listens (and sometimes responds) to that utterance. The process is repeated, with attention focused over a range of concepts, and utterances exchanged by both agents, until the agents form a shared lexicon that describes the range of concepts experienced by the agents.

The Lingodroids use language games to develop a shared lexicon to refer to places, distances, and directions based on the robots' cognitive maps. The robots play five types of language game: *where-are-we* is a game that generates names for places, which we call *toponyms*; *go-to* tests the toponyms by having one robot request a meeting at a toponym; *how-far* generates names for distances, referring to the offset between two locations; *what-direction* generates names for directions, referring to the angle between two remote locations relative to a third location; and *where-is-there* generates names for places that the robots have never visited by combining existing toponyms with distances and directions to refer to new places.

In this paper, we show that the Lingodroids can develop coherent symbols for places, distances, and directions, and can use those symbols to refer to novel places beyond the limit of their cognitive maps – imaginary places. By using *go-to* games to set meetings at the imaginary places, we show that the Lingodroids can effectively ground their knowledge of imaginary places to carry out a goal directed task. The major contributions of this paper are the demonstration of spatial language learning on real robots, and the first demonstration of real robots effectively grounding generative knowledge using an evolved language.

The paper proceeds with a brief review of related work before detailing the algorithms for building cognitive maps and playing language games. We describe the robots, the environment and the sequence and parameters used in the language games, and provide results that show the coherence of the lexicons, and the performance during the *go-to* tasks. The discussion focuses on the lessons learnt in cognition, and how those lessons will be applied in further studies.

This research was supported by Australian Research Council Discovery Grant DP0897078 and Special Research Initiative on Thinking Systems, TS0669699 to GW and JW and an Australian Postgraduate Award to AG.

R.S. and J.W. are with the School of ITEE at The University of Queensland, Brisbane, QLD, 4072, Australia, e-mail: ruth@itee.uq.edu.au.

A.G., M.M. and G.W. are with the School of Engineering Systems at the Queensland University of Technology, QLD, 4001, Australia. Tel: +61 7 3138 9974, e-mail: michael.milford@qut.edu.au

II. LITERATURE REVIEW

Spatial language forms the foundations of many aspects of human cognition and languages [7]. The development of grounded spatial languages for robots involves a range of challenges, some requiring development of cognitive systems and others relating to communication between robots [8-12]. For the robots to use language productively, they need to ground the words – that is, to link symbols and objects to their own sensors and behavior [13-15]. At its core, spatial language needs to be able to refer to places in the world, including places that cannot currently be perceived. Cognitive maps inspired by the hippocampus [3, 4] are learned directly through a robot's own experiences and hence can provide internal representations for directly linking spatial terms [8, 16].

Each cognitive map is private to the robot that developed it, and language games are a methodology for them to link their shared experiences, and hence develop a common lexicon [6]. A typical language game involves two robots engaging in a conversation about their current experience. To learn similar meanings for words, their attention needs to be directed to the same topic. A seminal project was the Talking Heads project, in which shared attention was achieved by pointing two cameras towards the same location in space and discussing the objects that could be seen at that location [17]. Through repeating language games many times, conversations cover a range of concepts enabling the development of a comprehensive lexicon shared by both robots. Shared attention is typically effortless for humans, but is challenging for robots which have no sense of themselves or the other robot [18]. Typically the behavior of the robots during a language game is used to ensure similar actions, coupled with proximity [19], sometimes aided by markers [20] or an audible handshake [21].

III. METHOD

A. RatSLAM

The RatSLAM system was used to perform spatial mapping and robot navigation [2]. RatSLAM is a robot SLAM system based on models of the mapping and navigation processes in the rodent brain. It consists of a number of components; a neural network of *pose cells* that perform filtering of data from the robot's sensors, a network of *local view cells* that encode distinct visual scenes, and the *experience map*, a graphical map that provides a semi-metric, topological environment representation that is used in navigation. The experience map is of particular relevance to the work described in this paper, as it provides the underlying spatial representation on which the lexicons are built and by which the robots navigate.

The experience map consists of nodes, called *experiences*, and links between those nodes that encode the experiences' relative spatial arrangement, and the time the robot took to transition between the connected experiences. To plan a path between the robot's current location and a goal location, the node in the experience map associated with the robot's current location is seeded with a time stamp value of zero. An iterative routine then propagates time stamp values to all other experiences. The shortest route to the goal is calculated by performing gradient descent from the goal location. Once a path is calculated, the robot navigates the path using a mixture of local movement behaviors and obstacle avoidance. A more detailed description of RatSLAM and the navigation processes is given in [2].

B. Lingodroids

The language abilities of the robots are provided by the Lingodroid system [11], which was first developed on simulated robots. In this paper we addressed the challenge of translating the Lingodroid system to real robots. Key issues in using real rather than simulated robots include an inherent ambiguity of local views, odometry, and shared attention. Imprecise and ambiguous sensorimotor systems can impact on the coherence of maps and utility of resulting languages. A system for forming maps and languages for real robots should be robust enough to deal with the imprecision and ambiguity of the real world. In the experiment described in this paper, the robots play location language games when they are within hearing distance, which is defined as both robots being able to hear a Dual-Tone Multi-Frequency (DTMF) beep emitted by the other robot. The audibility of the DTMF beep establishes the shared attention on the current location required for the language games.

In each language game, the hearer requests information from the speaker, who responds with an answer. The request and response utterances contain a sequence of toponyms, distance terms, and direction terms outlined in Table I, where x is the current toponym, y is the toponym describing the orientation of the robots, z is the target toponym, d is the distance word, and θ is the direction word. In all games except for the *go-to* game the response word may be invented by the speaker. At the end of all games, except for the *go-to* game, both speaker and hearer update their lexicon based on the speaker's response. After the hearing distance has been established with the DTMF beep, the utterances are transmitted wirelessly between the two robots, with each word comprising two consonant-vowel syllables.

TABLE I
LANGUAGE GAME UTTERANCES

Game	Request	Response
Where-are-we	-	X
Go-to	-	Z
How-far	x,z	D
What-direction	x,y,z	Θ
Where-is-there	x,y,d,θ	Z

In a *where-are-we* game, the speaker determines the best toponym for the current location, inventing a new one if no suitable word exists. Both the speaker and the hearer then update their lexicons based on the speakers' chosen toponym.

Go-to games were used to test the usefulness of the toponymic languages. In a *go-to* game, the speaker randomly chooses a toponym in its lexicon as the target. If both robots are able to plan a path to the target location, then a *go-to* game begins. Both robots then navigate to the target, and beep when they reach the target. Each robot records whether they found the target and heard the other robot. There are five possible outcomes of a *go-to* game: *both met*, when both robots hear each other at the target; *one met*, when both robots find the target but only one hears the other robot; *both found, not met*, when both robots find the target but neither hear each other; *one found*, when one robot reaches the target but the other does not; and *neither found*, when neither robot reaches the target.

After the robots have developed a toponymic lexicon, they are able to develop lexicons of distances and directions. In a *how-far* game, the hearer specifies two toponyms and the speaker determines the best distance concept to refer to the distance between the prototypes of the two toponyms. In a *what-direction* game, the hearer specifies three toponyms, corresponding to the 'current,' 'orientation,' and 'target' locations and the speaker determines the best direction concept to refer to the angle between their 'orientation' and the 'target' location when located at the 'current' location.

After the robots have developed lexicons for toponyms, distances, and directions, they are able to combine these concepts generatively to form new toponymic concepts. In a *where-is-there* game, the hearer specifies two toponyms, the 'current' and 'orientation' locations, a distance term, and a direction term. The speaker determines the referent location by this combination of concepts, and determines the best toponym to refer to that location, either choosing an existing word or inventing a new one. If the referent location does not correspond to an existing experience, the robot creates a pseudo-experience at that location. Pseudo-experiences are located in the same framework as the experience map, and are attached to nearby experiences so that updates to the experience map also update the pseudo-experiences.

The robots have three lexicon tables, one each for toponyms, distances, and directions. A lexicon table stores a count of the number of times each word has been used together with component parts of concepts, which we call *concept elements*. For toponyms, the concept elements are the experiences in the robot's map and pseudo-experiences created in *where-is-there* games. Distance and direction concept elements are created as the robots experience new distances and directions in the *how-far* and *what-direction* games. Distances are calculated from the distance between two locations in the robot's experience map. Directions are calculated from the angle described by the combination of three locations in the robot's experience map. In each game, the speaker determines the best word for the current concept element by calculating the confidence value for each word and choosing the word with the highest confidence value.

The confidence value, h_{ij}, for a word, w_j, at the concept element, s_i, was calculated as follows:

$$h_{ij} = \frac{\sum_{k=1}^{X} a_{kj}(D - d_{ki})/D}{\sum_{m=1}^{N} a_{mj}} \qquad (1)$$

where X was the number of concept elements within a neighborhood of size D of the current concept element, s_i; a_{ij} was the number of times that the concept element, s_i, and the word, w_j, had been used together; d_{ki} was the distance between concept element s_k and s_i; and N was the total number of concept elements that had been created by the robot. The concept element was specified by the request sent by the hearer. The neighborhood sizes used were 3m for the toponyms, 1.5m for distances, and 30° for directions.

In each interaction, words were invented with probability, p, as follows:

$$p = k \exp\left(\frac{-h_{ij}}{(1-h_{ij})T}\right) \qquad (2)$$

where $k = 1$, h_{ij} was the confidence value of the concept element-word combination, and T was the temperature, which effectively sets the invention rate for new words. Equation 2 allowed agents to use existing words when a word was associated with the current concept element with a high confidence, and to probabilistically invent words otherwise. Varying the temperature alters the word invention rate, where a higher temperature increases the probability of word invention. In the study described, the temperature decreased linearly from 0.1 to 0.0 during the experiment.

C. *Quality Measures*

Three measures were used to determine the quality of the lexicons developed by the robots: *production coherence*, *comprehension coherence*, and *go-to* game outcomes (described earlier). For the toponyms, production coherence and comprehension coherence were calculated using the unique experience maps of each robot rotated and shifted to find the best match between the two maps.

The production coherence of a lexicon was calculated over a set of concept elements. For toponyms, the set was the corners of a 0.25m grid, for distances, the set was every 0.25m, and for directions, the set was every 2.5°. The production coherence is the percentage of this set for which both robots produce the same word. Production coherence provides an indicator of whether the robots have a similar lexicon, with higher values indicating more similar lexicons.

The comprehension coherence of a lexicon was the average difference between the concept elements interpreted by the robots for each word in the lexicon, measured in meters for toponyms and distances, and in degrees for directions. Comprehension coherence provides an indicator of whether the robots construct similar concepts for each term in their lexicon, with a good value being less than the neighborhood size used for each concept type: 3m for toponyms, 1.5m for distances, and 30° for directions.

Good values for production and comprehension coherence indicate likely success in using the language, as measured by *go-to* games. Note that higher values are better for production coherence and lower values are better for comprehension coherence. A good value for *go-to* game success is well above the robots' chance of meeting each other at a randomly chosen location, which is dependent on the size of the room and the hearing distance of the robots.

IV. Experimental Setup

The experiment was conducted using two Pioneer3-DX platforms from Mobile Robots (see Figure 1). The Pioneers were equipped with a 360 degree panoramic camera rig, wheel encoders, a laser range finder and sonar for mapping and obstacle avoidance. A microphone and speakers were used for audible communication between Pioneers, tuned to a maximum hearing range of 1.5m. Communication was also performed through the wireless network for non-location-specific communication. All processing was performed on-board on a 2 GHz Pentium M processor.

Experiments were performed in an office environment consisting of a large central room with four smaller offices accessible on two sides (see Figure 2). Entry to each of the smaller offices could be controlled using the door. The lexicons were formed in this environment in five stages:

Initial Exploration: Both robots were confined to the central room to build individual maps. 100 *where-are-we* games were played to develop toponyms.

Initial Assessment: While restricted to the central room the toponym lexicon was assessed with 50 *go-to* games.

Generative Games: Without moving, 100 *how-far*, 100 *what-direction*, and 100 *where-is-there* games were played, with the robots developing a distance and direction lexicon, and adding to their toponym lexicon.

Further Exploration: The doors to the smaller offices were opened allowing the robots to individually expand their maps to include these areas. No language games were played during this stage.

Final Assessment: All toponyms were assessed with a further 50 *go-to* games.

V. Results

A. Initial Exploration

In the central room, the production coherence (the percentage of concepts for which both robots produced the same word) was 87.4% (see Figure 3). The comprehension coherence (the distance between word locations) was 0.65m, averaged over the positions of the five toponyms.

B. Initial Assessment

In the 50 *go-to* games played in this section, 38 games resulted in at least one of the robots hearing the other robot at the target, indicating that the robots found the target and were close to each other when they did (see Figure 4). In one game both robots found the target but neither heard the other agent. In the remaining eleven games, only one robot found the target. This occurred only for the toponyms "pize", "reya", and "rije", which were all located in tight corridors in the room. In these locations, if one robot stopped at their target, the other robot was sometimes unable to reach their target on the other side of the stationary robot. The calculated chance of meeting for the two robots, if both chose a random location as the target, is 9.2%, with a hearing distance of 1.5m and a room area of 76.5m^2.

Fig. 1. A successful *go-to* game. Both Pioneers meet at toponym *kuzo* after it has been grounded through *where-are-we* games.

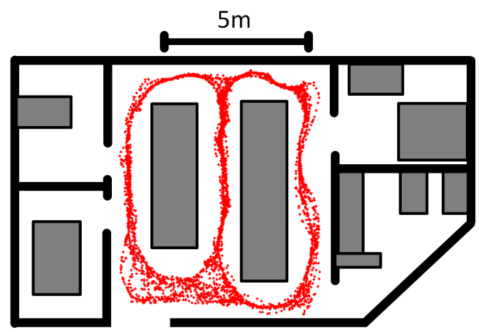

Fig. 2. The world of the robots superimposed on the experience map of one of the robots after the initial exploration stage. The experience map has been rotated and scaled to fit the floor-plan.

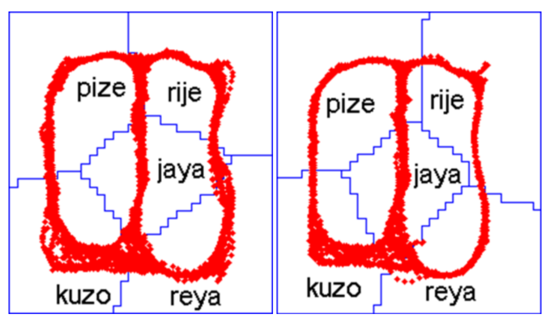

Fig. 3. Experience maps and toponym lexicon developed during initial exploration for each robot.

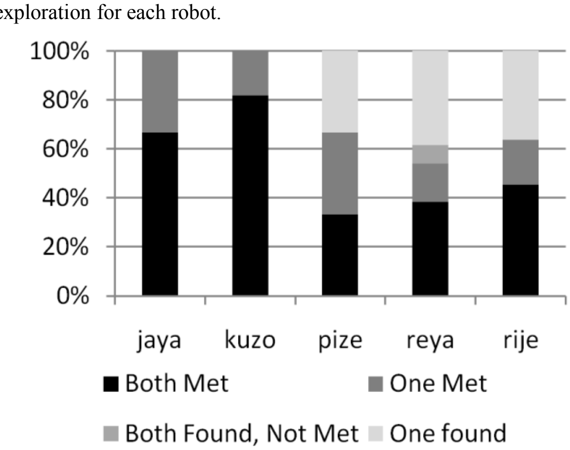

Fig. 4. *Go-to* game outcomes for each toponym in the initial assessment.

This stage demonstrates that the development of a toponymic language through *where-are-we* games results in the ability for the robots to specify a set of locations accurately through shared symbols. The breakdown in the successful outcomes of *go-to* games comes with the behavioral challenge in the co-location of two robots.

C. Generative Games

The generative games provide the ability to refer to a novel location with a shared symbol. In this stage, the agents developed distance and direction terms that enabled them to ground generative references to locations. The robots developed four distance terms with a production coherence of 73.2% and a comprehension coherence of 0.375m (see Figure 5). Five direction terms were developed with a production coherence of 89.7% and a comprehension coherence of 10.0°. At the end of the 100 *where-is-there* games, the robots had 30 words for toponyms, including the five of the original lexicon, with a production coherence of 37.2% and a comprehension coherence of 3.26m (see Figure 6). The considerable decrease in coherence compared to stage 1 is the result of the uncertainty (characterized by both production and comprehension coherence) that is inherent in the symbols used to generate the new toponyms.

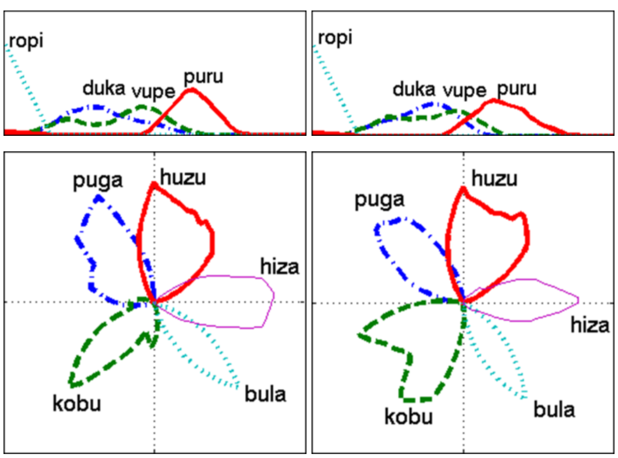

Fig. 5. Distance and direction lexicons developed during the generative games for each robot.

D. Further Exploration

In the further exploration section, the robots did not interact with each other, but further explored their world with the doors to the smaller offices open. The result of the exploration was that some toponyms developed during the previous section that referred to locations beyond the edge of their experience maps now refer to locations within the borders of their experience maps (see Figure 6).

E. Final Assessment

In the 50 *go-to* games played in this final section, 31 were to locations that were not part of the initial toponymic lexicon. Of the 31 games, 16 resulted in at least one of the robots hearing the other robot at the target (see Figure 7). In six games both robots found the target but neither heard the other agent. In seven games only one robot found the target, and in two of the games neither was able to find the target before a time-out. These results indicate that the communication of the 'imagined' concepts was not perfect, but certainly improved success at goal directed behavior (52%) to well above chance (9.2%).

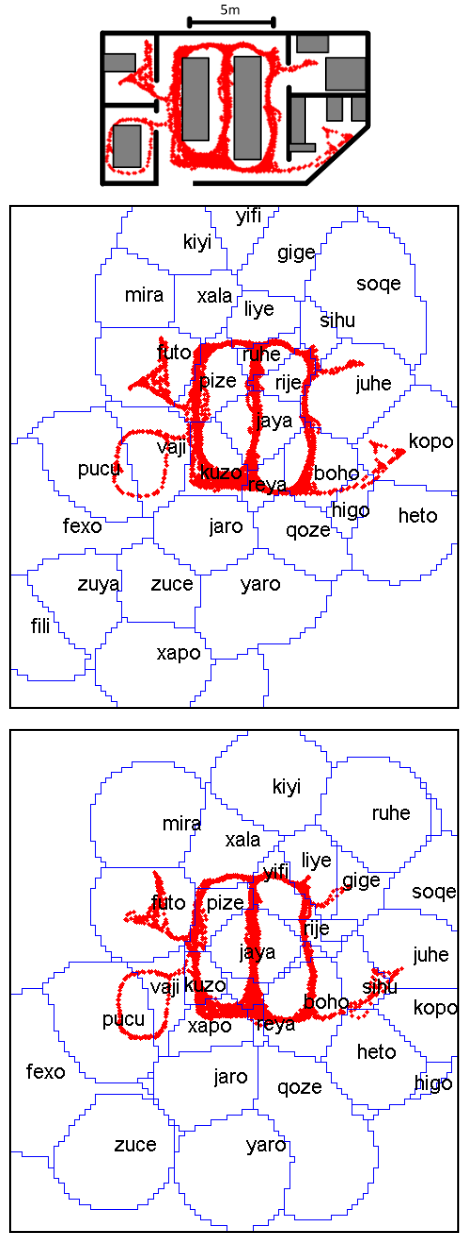

Fig. 6. Extended toponym language developed during the generative games overlaid with the experience map after further exploration for each robot. The floor-plan with superimposed experience map is provided as a reference.

The outcome of the *go-to* games is best described using specific examples. The meeting room, to the bottom left of the map, had two words, with one referring to either side of the room: *pucu* and *vaji*. The robots played a total of nine games using these two words, of which seven were successful with at least one robot hearing the other at the target location, and two games in which one robot was unable to find the target location within four minutes. An interesting example of where the robots failed to meet each other is *juhe*. The area covered by the word spans the area between the two rooms on the right side for both robots, but the location is interpreted by one robot as the top room and the other robot as the bottom room, making it impossible for the two robots to meet each other at *juhe*.

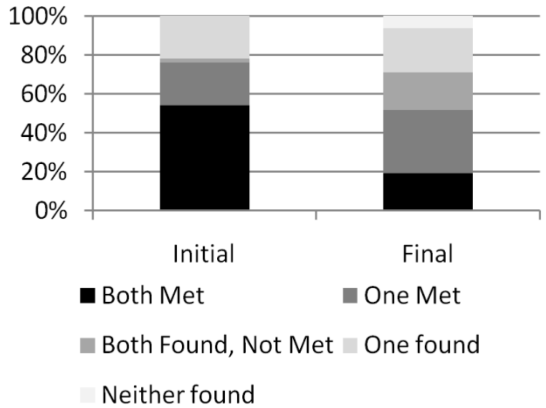

Fig. 7. *Go-to* game outcomes for the initial and final assessment.

VI. Discussion

Grounding of generatively developed or 'imagined' concepts is necessary to be able to discuss anything beyond direct or remembered experience. For those concepts to be considered truly grounded, they must be able to be used in sensible ways.

The experiment described in this paper demonstrated success at the meeting task set for the robots. The robots formed independent maps of their world, grounded in personal experience, and developed a shared toponymic language, grounded in shared experience. They formed distance, direction, and additional toponyms through generative processes. The final assessment showed that these generatively grounded toponyms had real grounded meaning, with the robots successfully navigating to and meeting at several of the shared, imagined toponyms. The experiment extended our previous work [11] through the implementation of the system on real rather than simulated robots, and was the first time that the robots played *go-to* games to meet at imagined toponyms.

This work provides a platform to study the cognitive processes involved in knowledge representation, planning, language development, symbol grounding, and imagination.

To extend the current work, behavioral and environmental challenges need to be addressed. If the robots had more flexible ways of interacting with each other, for example one robot asking the other to move out of the way so that it can reach the target, then *go-to* games in tight spaces would be more successful. An alternative to the autonomous acquisition of language through language games is to provide robots with algorithms for extracting meaning from natural language. This technique has been used successfully to form a variety of concepts, including those required for following navigational directions [9, 10]. Insights from these studies may provide additional ways for directing shared attention to allow more interesting concepts to form, such as descriptions of how to get to a place or the accessibility of places beyond the edges of the current map.

References

[1] M. Milford, G. Wyeth, and D. Prasser, "Simultaneous localisation and mapping from natural landmarks using RatSLAM," in *Australasian Conference on Robotics and Automation* Canberra, 2004.

[2] M. Milford and G. Wyeth, "Persistent Navigation and Mapping using a Biologically Inspired SLAM System," *The International Journal of Robotics Research,* vol. 29, pp. 1131-1153, August 1 2010.

[3] J. O'Keefe and L. Nadel, *The hippocampus as a cognitive map*. New York: Oxford University Press Inc, 1978.

[4] M. Milford, G. Wyeth, and D. Prasser, "RatSLAM: a hippocampal model for simultaneous localization and mapping," in *IEEE International Conference on Robotics and Automation, ICRA 2004,* New Orleans, LA, USA: IEEE Press, 2004, pp. 403-408.

[5] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena,* vol. 42, pp. 335-346, 1990.

[6] L. Steels, "Language games for autonomous robots," *IEEE Intelligent systems,* vol. 16, pp. 16-22, 2001.

[7] S. C. Levinson, *Space in language and cognition: Explorations in cognitive diversity*. Cambridge, UK: Cambridge University Press, 2003.

[8] M. Levit and D. Roy, "Interpretation of Spatial Language in a Map Navigation Task," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 37, pp. 667-679, 2007.

[9] C. Matuszek, D. Fox, and K. Koscher, "Following directions using statistical machine translation," in *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction* Osaka, Japan: Association for Computing Machinery, 2010.

[10] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions," in *Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction* Osaka, Japan: Association for Computing Machinery, 2010.

[11] R. Schulz, D. Prasser, P. Stockwell, G. Wyeth, and J. Wiles, "The formation, generative power, and evolution of toponyms: Grounding a spatial vocabulary in a cognitive map," in *The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)*, A. D. M. Smith, K. Smith, and R. Ferrer i Cancho, Eds. Singapore: World Scientific Publishing, 2008, pp. 267-274.

[12] R. Moratz and T. Tenbrink, "Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations," *Spatial Cognition & Computation,* vol. 6, pp. 63-107, 2006.

[13] N. Mavridis and D. Roy, "Grounded Situation Models for Robots: Where words and percepts meet," in *International Conference on Intelligent Robots and Systems*, 2006, pp. 4690-4697.

[14] D. Roy, "Semiotic Schemas: A framework for grounding language in action and perception," *Artificial Intelligence,* vol. 167, pp. 170-205, 2005.

[15] L. Steels, "The symbol grounding problem has been solved. So what's next?," in *Symbols and Embodiment: Debates on meaning and cognition*, M. De Vega, A. M. Glenberg, and A. C. Graesser, Eds. New York: Oxford University Press, 2008, pp. 223-244.

[16] M. Milford, R. Schulz, D. Prasser, G. Wyeth, and J. Wiles, "Learning spatial concepts from RatSLAM representations," *Robotics and Autonomous Systems - From Sensors to Human Spatial Concepts,* vol. 55, pp. 403-410, 2007/5/31 2007.

[17] L. Steels, *The Talking Heads Experiment* vol. I. Words and Meanings. Brussels: Best of Publishing, 1999.

[18] A. D. M. Smith, "Establishing communication systems without explicit meaning transmission," in *ECAL01*, J. Kelemen and P. Sosik, Eds. Prague: Springer, 2001, pp. 381-390.

[19] L. Steels and J.-C. Baillie, "Shared grounding of event descriptions by autonomous robots," *Robotics and Autonomous Systems,* vol. 43, pp. 163-173, 2003.

[20] P. Vogt, "Bootstrapping grounded symbols by minimal autonomous robots," *Evolution of Communication,* vol. 4, pp. 87-116, 2000.

[21] L. Steels and M. Loetzsch, "Perspective alignment in spatial language," in *Spatial Language and Dialogue*, K. R. Coventry, T. Tenbrink, and J. A. Bateman, Eds. Oxford, UK: Oxford University Press, 2007, pp. 70-88.